# Formation of Stars and Planetary Systems

Prof. dr. Alex DE KOTER

Astronomical Institute *Anton Pannekoek*
University of Amsterdam
P.O. Box 94249, 1090 GE Amsterdam,
The Netherlands

Instituut voor Sterrenkunde,
KU Leuven
Celestijnenlaan 200D, 3001, Leuven,
Belgium

Version 1.48.00

# Contents

# 1

# Introduction

The complex of processes known as star formation must have occurred innumerable times in the past. The Big Bang, after all, did not produce a universe full of stars but of diffuse gas. How gas turns into stars, and how, as a by-product, planets are formed, is the subject of this lecture series. Anyone wishing to study the problem is aided immeasurably by the fact that star formation is also occurring now, and in regions close enough that the transformation from gas to stars can be examined in some detail.

## 1.1 Basic observational hints to stellar origins

The structure and evolution of a star can be seen as an adaptation to the self-gravity of a considerable amount of gas. Self-gravity is compensated by a pressure gradient, associated with a density and temperature gradient. For stars the latter is crucially important and is supported by energy transport, the energy being produced by nuclear fusion reactions in the stellar interior. As the energy source is not infinite, stars do not have eternal life. The fact that stars exist, obviously implies that they have originated sometime, somewhere and somehow.

Stellar lifetimes critically depend on the mass of a star, massive stars having a much shorter life expectancy than lower-mass ones. The reason is that higher mass implies higher central temperatures and that the nuclear reaction rates increase strongly with temperature. The nuclei which fuse in stars have velocities that belong to the high-energy tail of the Maxwell distribution, hence their amount depends exponentially on the temperature. The life expectancy of the Sun is some 12 Gyr (of which it will spend some 10 Gyr on the main sequence; Turcotte et al. 1998) or some 50 galactic orbits; that of a 10 $M_\odot$ star is less than 25 Myr, a 'galactic month' (see section 1.2). Massive stars thus hardly move away from the environment where they are born in the global galactic context, and the observation that essentially all high-mass stars in the Milky Way are located in the galactic plane then strongly suggests that present-day star formation must occur in the plane.

Massive stars tend to be grouped in clusters and associations. It appears then likely that most,
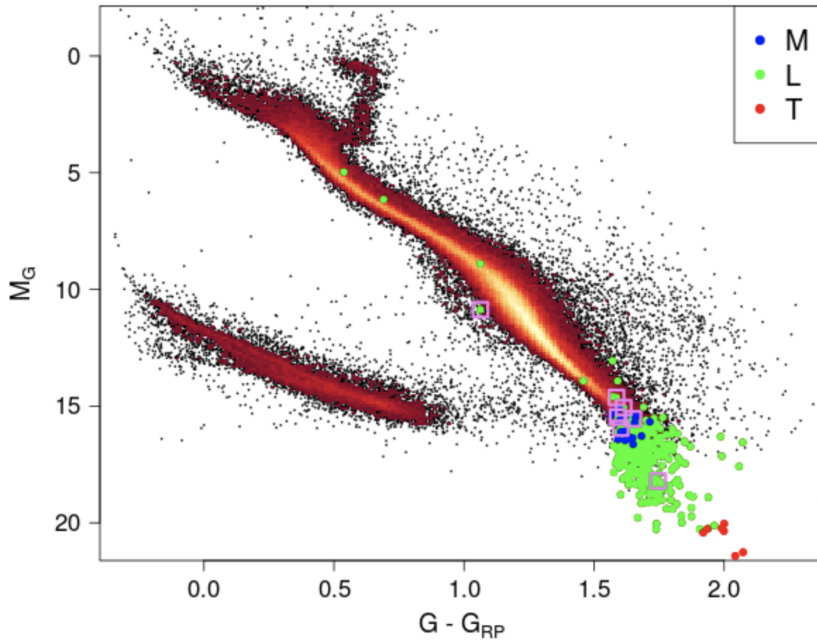
Figure 1.1: *HRD of 240 703 stars with parallax $\varpi > 10$ (or distance $d < 100$ pc) observed with GAIA. Cross matches with the Gaia Ultracool Dwarf Sample (GUCDS; Smart et al. 2017) of brown dwarfs and gas giants are given the color blue (M-type object), green (L type) or red (T type). The magnitudes of (in any case) four green objects are too bright, most probably because of a cross-match issue. This leaves 21, 443, and 7 of M-, L, and T-type objects, respectively. Pink squares are added around stars with tangential velocity $v_\mathrm{T} > 200\,\mathrm{km\,s^{-1}}$. From: Gaia Collaboration 2018).*

if not all, stars are born in large groups, from which mature stars may eventually escape. The (more or less) coeval formation of stars in clusters is a most beneficial circumstance and helps to further our understanding of star formation at a galactic scale, as the color-magnitude diagrams of clusters allow fairly reliable age determinations. Open clusters where the most massive stars are gone, having ended their lives in core-collapse supernovae, still are quite well concentrated in the plane of our galaxy; several clusters with ages more than 1 Gyr are known.

It consistently appears that the youngest clusters are embedded in a gaseous and dusty medium. This suggests that obvious stellar nurseries are then the densest phases of the interstellar medium, the so-called molecular clouds. In many such clouds the evidence for star formation indeed is high: they contain young stars, and entities which clearly are good candidates for pre-stellar objects. The observational study of star formation hence strongly relies on that of molecular clouds.

The tools at hand are diverse, and cover the full electromagnetic spectrum. Since the media themselves are typically cold and obscuring, the infrared, sub-millimeter and radio regimes are of obvious relevance, both to image the internal structure of the clouds and to exploit the
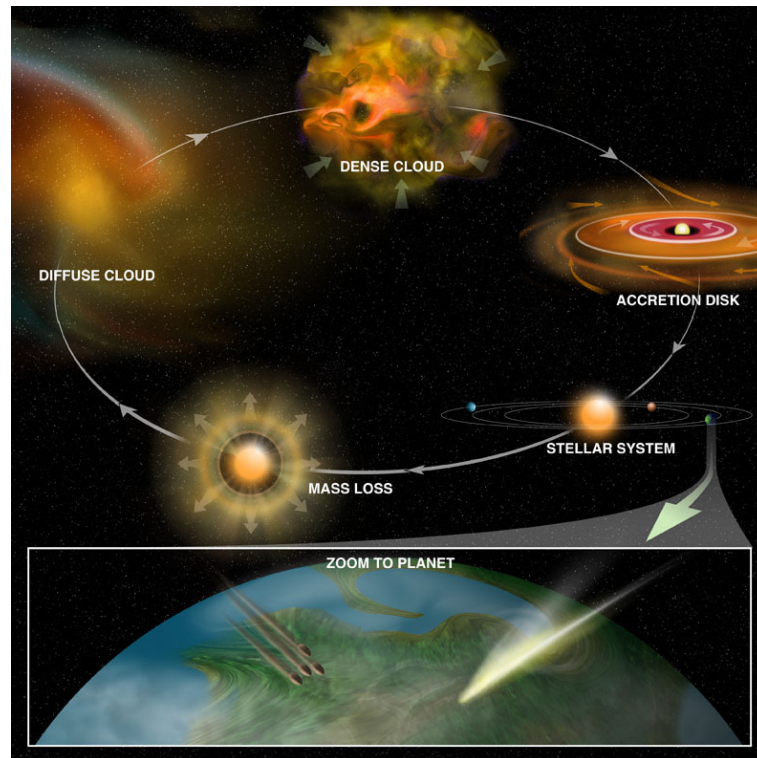
Figure 1.2: *A sketch of the stellar life-cycle. Stars form when dense molecular cores form in the diffuse gas in the interstellar medium. Because of angular momentum conservation, the forming stars accrete their matter through infall via a circum-stellar disk. In these disks a planetary system may form. Particularly near the end of the lives of stars they eject their envelopes in a stellar wind. This mass loss brings products of thermo-nuclear fusion into the ISM, which mixes with interstellar gas. In time, a new cycle of star-formation may start in molecular clouds that have been enriched with these new elements. From: Bill Saxton NRAO/AUI/NSF.*

rich spectral diagnostics from molecules and dust species. However, star formation involves a rather violent conversion of energy, and the X-ray, ultraviolet and optical windows are also relevant, to study the accretion of gas onto (proto-)stars and shocked outflows from (proto-)stars in all mass ranges, and also the effects of radiation-driven winds from young massive stars on the (molecular) cloud material.

## 1.2 The stellar life-cycle

Figure 1.1 shows an Hertzsprung-Russel diagram (HRD), with the positions of almost a quarter of a million stars within 100 pc distance. Stars spend the majority of their life-times as a main-sequence star, during which time its energy source is the fusion of hydrogen to he-

lium deep in the core of the star. Main-sequence stars obey a mass-luminosity relation, with $L \propto M^\eta$. The slope $\eta$ changes slightly over the range of masses; between 1 and 10 $M_\odot$, $\eta \simeq 3.66$. The relation flattens out at higher masses, due to the contribution of radiation pressure in the central core. This helps to support the star, and decreases the central temperature slightly. The relation also flattens significantly at the very faint end of the luminosity function. This is due to the increasing importance of convection for stellar structure.

With an estimated main-sequence life of the sun of 10 Gyr, we may estimate $\tau_{\mathrm{MS}}$ for other stars using the relation

$$\tau_{\mathrm{MS}} = 10 \, (M/M_\odot)/(L/L_\odot) \, \mathrm{Gyr} = 10 \, (M/M_\odot)^{-2.66} \, \mathrm{Gyr}. \tag{1.1}$$

For a 10 $M_\odot$ star we find 22 Myr. All such main-sequence stars in the sky therefore can at most be this old.

Later in their lives, when the majority of the hydrogen in their cores has been processed by fusion into helium, stars leave the main sequence, expanding to become giants or supergiants, and undergoing various stages of losing mass. This ejected material can then form some of the ingredients for subsequent generations of stars.

Hence, we see that stars undergo a life-cycle in which new generations of stars are formed in part from the debris of previous generations of stars (see Fig. 1.2). In fact almost all the material which forms a new generation of stars, apart from hydrogen and helium, is the product of fusion in the centers of stars of previous generations. Thus the formation of stars is an integral part of a cyclic process or stellar life-cycle. These lecture notes cover part of this stellar life-cycle, starting at the birth-place of stars and following the progress from there all the way to the point at which stars join the main sequence.

## 1.3 Star formation at a galactic and cosmic scale

In the 1970s it was recognized that some galaxies experience an exceptionally high rate of star formation. In such a star-bursting galaxy, the rate of star formation is so large that the galaxy will consume all of its gas reservoir, from which the stars are forming, on a timescale much shorter than the age of the galaxy. As such, the starburst nature of a galaxy is a transient phase in its evolution. The majority of starburst galaxies are caused by galaxy-galaxy interaction, inducing tidal streams and potentially leading to a merger of the two systems. Well-known starburst galaxies include M82, the Antennae Galaxies (NGC 4038 and NGC 4039), and IC 10.

The bottom image of figure 1.3 shows the starburst galaxy M82, about 3.2 Mpc away. Because of its unusual appearance it was once thought that this galaxy was exploding. The current interpretation is that between 10 and 100 Myr ago the inner 400 pc of M82 began a tremendous episode of star formation that is still continuing. The center of M82 contains a wealth of O and B stars and supernova remnants that have expelled more than $10^7 \, M_\odot$ of gas from the plane of

Figure 1.3: *The galaxies M81 (top left) and M82 (top right and bottom) at a distance of 3.2 Mpc. The image of the starburst galaxy M82 in the bottom panel is somewhat rotated relative to the top panel image. The composite image was obtained by the Subaru Telescope's FOCAS instrument in $B$, $V$, and $H\alpha$. From: National Astronomical Observatory of Japan (NAOJ).*

the galaxy. All of this violent activity may have been triggered by a tidal interaction with M81, a spiral galaxy that is now about 36 kpc away from M82 (as projected in the plane of the sky; see the top image of figure 1.3) and linked to it by a gaseous bridge of neutral hydrogen. In fact also M81 is rich in star formation, with many star clusters and associations being formed in its spiral arms. The center of M81 is known to be a significant source of X-rays. This has been interpreted as implying that M81 has a central black hole that is being 'fed' by in-falling gas[1]. This gas infall may also be a consequence of the tidal interaction with M82.

---

[1]The mass of the central black hole in M81 has been estimated to be $\sim 70$ million solar masses; for comparison, the black hole in our galaxy weighs in at $\sim 3$ million solar masses.
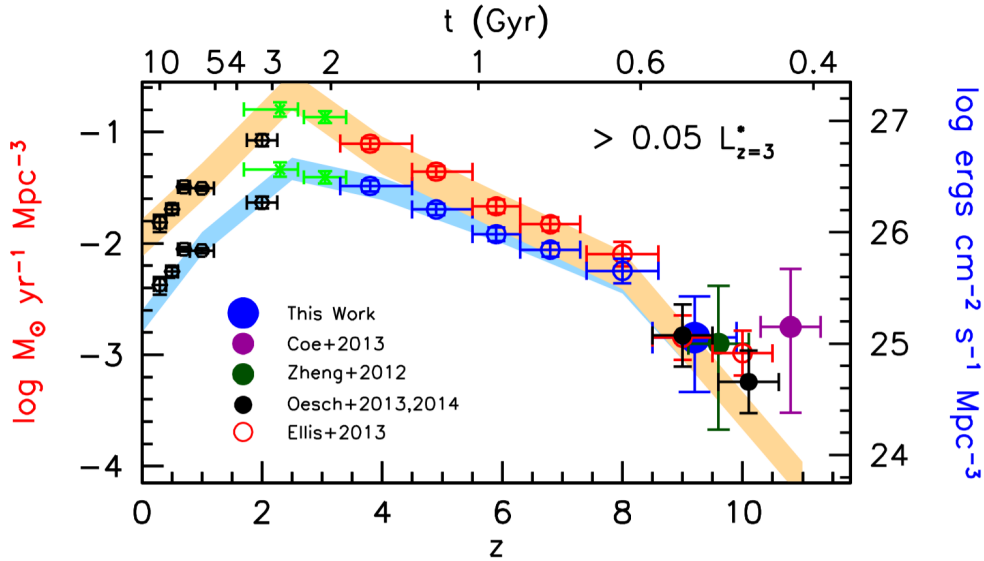
Figure 1.4: *The rest-frame continuum ultraviolet luminosity density (right axis, blue bar and blue points), and the star formation rate density (left axis, red points and orange bar) throughout 13.2 Gyr of cosmic history. The points at a redshift of about 10 show that galaxies existed at a time when the universe was only 500 Myr old. The star formation rate density is much smaller ($\sim 10$ percent) at this time than it is just $\sim 200$ Myr later at $z \sim 8$. This demonstrates how rapid galaxy build-up was at $z \sim 10$, showing the epoch from $z \sim 10$ to 8 to be a crucial phase in the assembly of the earliest galaxies. At about a redshift of 3-2, equivalent to a cosmic age of 2 to 3 Gyr, the universe most actively formed stars. Approximately half the stars in the present-day universe were in place at that time. Since that time, the star-formation rate has declined by about an order of magnitude, painfully illustrating that the heydays of our universe have passed. To emphasize this a bit more: at $z \sim 0.7$, nearly 40 percent of intermediate and high-mass galaxies are actively starbursting. At $z \sim 0$ only 1 percent. From: Bouwens et al. (2014).*

Throughout cosmological time, clustering and subsequent merging has been an important mechanism by which galaxies gain mass. In addition, it governs their shape and structure. Observations show that indeed many – but not all – of the galaxies present at a cosmic age of $\sim 1$ Gyr had sizes and masses of only about a percent of the size and mass of the present-day Milky Way. The strong link between galactic-scale star-formation and galaxy merging implies that by tracing star formation back to the earliest times one is directly probing cosmic structure formation and galaxy evolution. Figure 1.4 shows this star-formation history of the universe as derived from tracing the ultraviolet light of galaxies, i.e. light produced by young massive stars, and converting this – correcting for extinction of UV light by dust where needed and the presence of newly formed stars that do not emit copious amounts of UV light – to a total mass of newborn stars. The figure displays that the peak in star-forming activity of our universe occurred some 2-3 Gyr after the Big Bang, and that since that time the star-forming activity has declined by about a factor of ten. Indeed, it seems that in this sense the universe is passed its peak.

## 1.4 First Stars and First Binaries

Figure 1.4 also shows that the first galaxies already existed at a cosmic age of $\sim$500 Myr. Therefore, the First Stars and First Binaries likely formed several hundreds of millions of years after the start of time. The properties of the first stars are anticipated to have had an important impact on the early evolution of the Universe, as the ultraviolet photons they produce destroy $H_2$ molecules and ionize hydrogen, and their stellar winds and supernovae first enriched the interstellar medium with heavy elements. The latter substantially affects the formation of the later generations of stars. If, as seems likely, the first stars were very massive (see chapter 8.3), they also could have contributed to black-hole formation in the early Universe, perhaps supplying the seed black-holes that – through galaxy merging and associated black-hole merging – grew into the supermassive central black holes of galaxies like the Milky Way and M81 in the present universe.

How did these first stars form? What were their birth properties? It is clear that conditions then were very different from those in the Galaxy at the present time, identifying this topic as a frontier of astronomy and cosmology (see chapter 8).

## 1.5 By-products of star formation: planets

Angular momentum conservation during the collapse phase of star formation predicts that stars are born with equatorial disks. Observations confirm that such disks are fairly widespread. They tend to disappear from our view on a time scale of some 10 Myr. A tantalizing possibility is that the gas and dust of the disks is not completely removed from the system, but (partly) settles into a limited amount of larger bodies, in doing so decreasing the total radiating surface of the disk. Indeed, in our own planetary system a large fraction of the dust mass is locked into a few planets, which orbit the sun in planes that are only slightly inclined with respect to the equatorial plane of the sun. Is planet formation in proto-planetary disks around young stellar objects a fairly natural by-product of star formation?

Indeed, this is what for instance NASA's Kepler space mission has revealed (see figure 1.5). Using planetary transits to detect planets orbiting their host star, and bias correction techniques to correct for the planets in systems that we do not observe at inclinations leading to observable transits, it was found that about 77% of stars with effective temperatures $T_{\rm eff} < 5000$ K have planets and that the average number of Kepler-like planets these stars have is at least 2.8. These fractions will likely increase when planets with periods longer than 400 days and radii less than about $0.4\,R_\oplus$ are added to the count.

Before the first exoplanets were known, reasonable agreement existed in views on how the planets in our solar system took shape, starting with coagulation of dust particles into larger bodies, some of which eventually grew massive enough to accrete gas from the disk as well. It
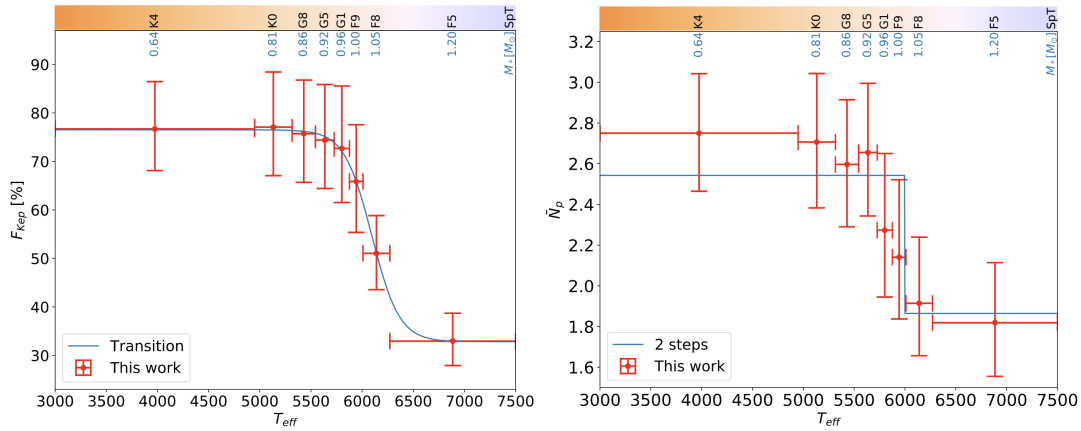
Figure 1.5: *Results from the final Kepler Data Release (DR25) catalog, showing the fraction of stars with planets $F_{\mathrm{Kep}}$ and the number of planets per planetary system (i.e., planet multiplicit) $N_p$ as a function of stellar effective temperature (as well as stellar mass). These numbers have been corrected for three types of biases: transit geometry bias, detection efficiency bias, and vetting efficiency bias. Only Kepler-planets with orbital period $P < 400\,\mathrm{d}$ and planet radius larger than about $0.4 R_\oplus$ are considered here because the Kepler detection efficiency rapidly drops at larger periods and smaller planet sizes. The circles and error bars indicate $50 \pm 34.1$ percentiles from the posterior distributions after Markov Chain Monte Carlo fitting. The blue curve shows the best fit of the model chosen to represent the data. At the top, the median mass and the corresponding spectral type is given using the spectral-temperature relationship as in Pecaut & Mamajek (2013). Indeed, planets appear to be ubiquitous: about 77% of late-type stars ($T_{\mathrm{eff}} < 5000\,\mathrm{K}$) have Kepler-like planets with an average planet multiplicity of $\sim 2.8$. From: Yang et al. 2020.*

was thought that gas accretion could only occur beyond the snow line in the disk, because only in this region would there be enough solid material available to grow planet cores of sufficient mass to accrete the gas as well (see figure 1.6). The detection of Jupiter-sized planets much closer to their parent star has challenged the basic picture of planet formation. A variant of the basic picture of planet formation is that giant planets are always formed beyond the snow line, but that viscous and tidal interactions with the gas in the disk can cause important migration of these planets. An alternative model states that the basic picture is wrong, and that (exo-) Jupiters owe their existence to processes not unlike those which form stars in the first place, i.e. gravitational instabilities in the circum-stellar disks.

The evolution of disks and that of planets inside them is a recent and currently extremely active field of study, both theoretically and observationally. It is fair to say, however, that in the present developments observations tend to be ahead of theory. Indeed, the observational tools have improved considerably over the recent decades, in sensitivity, spectral coverage, and spatial resolution, and they have revealed an unanticipated diversity of phenomena which it will take some time to fully understand.
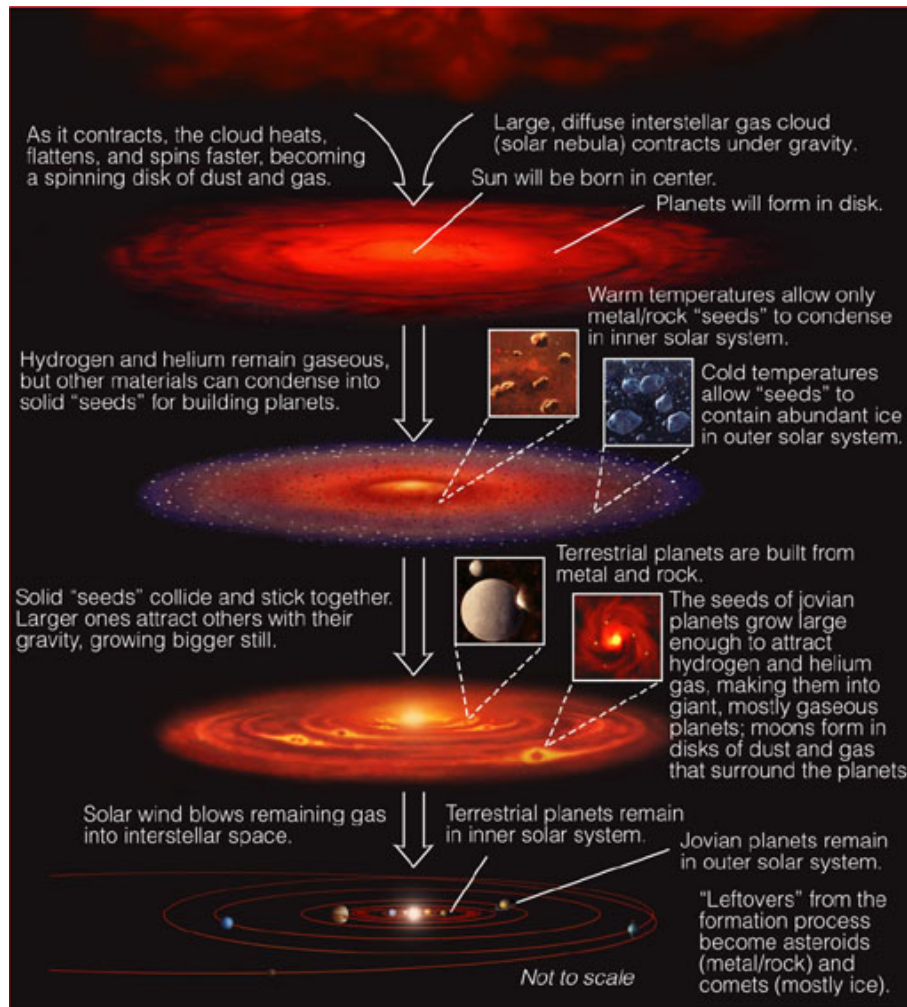
Figure 1.6: *Schematic overview of the formation of the planets in our solar system, distinguishing between inner terrestrial planets and the outer gaseous planets. From: Copyright 2004 Pearson Eduction, publishing as Addison Wesley.*

## 1.6  Objectives of star-formation and planet-formation theory

The overall goal of star-formation research is to develop a complete theory for the whole process. However, we are far from this overall goal, so it is appropriate to define some more specific objectives, in the form of questions. One should realize, however, that none of these questions has a definitive answer. This field is one of continuing research and discovery.

- 1 - What causes the initial mass function (IMF) for star formation? One interesting aspect of the IMF is that it appears to differ very little from one region to another, and therefore

it may be essentially universal.

- 2 - What determines the initial statistics of binary systems? By this we mean the initial binary fraction and the distributions of mass ratio, period, and eccentricity as a function of the mass of the primary star. Must all stars have a companion, be it a star, brown dwarf or giant planet, to deal with the angular momentum problem in collapsing clouds?

- 3 - Which fraction of stars form planetary systems and how is the diversity of planetary systems that are now being discovered explained (in a statistical way)? Is there a typical planetary mass, composition and/or typical planetary distance? More general, what is the initial mass function for planets and what remains of this function after dynamical instabilities cause bodies to collide or be ejected out of the system?

- 4 - A related question is what fraction of planetary systems include Earth-like planets and how many of these are in the zone in which water can exist in the fluid phase? The latter is thought to be an essential requirement to create an environment that is suitable for life to form and be sustained[2]. In how many life-friendly environments did life actually develop and proliferate?

- 5 - How do all aspects of star formation depend on environmental factors? These may vary enormously with cosmological epoch. By 'environmental factors' we mean external factors such as: mergers of galaxies and close interactions between galaxies; the strength of the shock waves associated with spiral modes in galaxies, and the depth of the galactic potential well in which the star formation occurs; the level of turbulence in the interstellar medium, and the strength of the interstellar magnetic field; the ambient intensity of high-energy ionizing radiation such as ultraviolet and X-ray photons, and cosmic rays; the temperature of the cosmic microwave background radiation field; and the heavy-element abundance in the interstellar gas.

- 6 - How did the first stars form? Did the first stars have companions, perhaps also gaseous planets? How massive were the first stars? How did star formation in the early history of the Universe differ from today's star formation? Conditions in the early history of the Universe were very different from what they are now. Two of the major ingredients which favour present-day star formation – low temperature and the presence of about 1 percent by mass of material in the form of dust grains – simply were not present. Also, magnetic fields were minimal, and less triggers stimulating star formation (such as stellar wind induced shocks or supernovae in the present-day universe) were present.

- 7 - How do massive stars form in the present-day Universe? Do they form through core-accretion as do their lower mass cousins or is an altogether fundamentally different mechanism required – such as the merging of multiple lower mass stars? Spherically accreting protostars at some point become so luminous that strong radiation pressure prevents further accretion, quenching growth and limiting the mass to several tens of

---

[2]From a cosmic perspective therefore, the formation of life may be seen as a by-product of a by-product of star formation.

solar masses. But several hundred solar mass stars have been found. How are these assembled?

## 1.7 Literature

No books fully cover what is presented in these lecture notes. However, we follow parts of:

- Stahler (Steven W.) and Palla (Francesco): *The Formation of Stars*, 1st Edition 2004 (2nd Reprint 2011). Wiley-VCH Verlag GmbH & Co, KGaA. Excellent standard book, focussing on star formation. Threats part of the basic physics in more detail than is done in these lecture notes.

- Bodenheimer (Peter H.): *Principles of Star Formation*, 1st Edition. Springer-Verlag Berlin Heidelberg 2011, as part of the series Astronomy and Astrophysics Library. Moderns book, solely treating star formation. Also discusses SPH simulations, Massive star formation and formation of the First Stars.

- Ward-Thompson (Derek) and Whitworth (Anthony P.): *An Introduction to Star Formation*, 1st Edition. Cambridge University Press 2011. Excellent and easy to read book focussing on concepts of star formation.

- Maeder (André): *Physics, Formation and Evolution of Rotating Stars*, 1st Edition 2009. Springer-Verlag Berlin Heidelberg 2009, as part of the series Astronomy and Astrophysics Library. Modern book, that discusses the effects of rotation in star formation; massive star formation, and the formation of First Stars.

- Armitage (Philip J.): *Astrophysics of Planet Formation*, 1st Edition. Cambridge University Press 2010. Outstanding book treating the basic physics of planet formation. The chapters on planet formation in these lecture notes closely follow this book.

- Kamp (Inga): *Star and Planet Formation*: Lecture notes of the University of Groningen. Used as a guide in writing several chapters of these lecture notes. Treats the chemistry of disks in some detail.

- Dominik (Carsten): *Star and Planet Formation*: Powerpoint lecture slides of the University of Amsterdam. Clear and concise notes treating most of the physics discussed in these lecture notes.

- Ormel (Chris): *Planet Formation*: Lecture notes of the University of Amsterdam. Clear and concise notes that focus on planet assembly and planet dynamics.

## Exercise 1.1

The minimum mass of O-type stars is $16\,M_\odot$. That of a B-type star is $3\,M_\odot$. The minimum mass of massive stars, i.e. stars that in the end experiences a core-collapse supernova explosion is likely in the range of $8$–$12\,M_\odot$ according to Poelarends et al. (2008), who propose a fiducial value of $9\,M_\odot$. What is the maximum age that an O-type or B-type star can reach? A supernova remnant is visible for about $50\,000$ yr. How long after a co-eval starburst does the last supernova fade away?

## Exercise 1.2

Figure 1.1 shows the HRD diagram for stars, brown dwarfs, gas giants, and white dwarfs observed by GAIA as isolated objects, all within a distance of $100$ pc. Note the red dots, which represent gas giants.

a) Calculate the star (read: object) density in the volume probed in $\#\,/\,\mathrm{pc}^3$.

a) Will GAIA probe the full population of giant planets plotted in this diagram?

b) Looking at the diagram, ponder a bit on the possibility of different modes of planet formation? How many modes of planet formation do you distinguish (with your present knowledge)?

## Exercise 1.3

The Kepler space mission used continuous-monitoring photometry to search for planetary transits in a $115\,\mathrm{deg}^2$ field of view (FOV) for over 9 years. Kepler observed $530\,506$ stars and detected $2\,662$ planets. The observed statistical properties of these planets are however not unbiased. Think of at least two reasons why the observed properties may suffer from biases – i.e., not be representative of the intrinsic distribution of planet properties in the FOV – and in what way. Think of at least one reason why the observed statistical properties may not be universal – i.e., not be representative for every location in space.

## Exercise 1.4

Study Gillon et al., Nature 542, 456 (2017), and summarise the essential results in a few short paragraphs.

# 2

# Molecular Clouds

## 2.1 Different phases of the interstellar medium

'The Interstellar Medium is anything not in stars' (Donald Osterbrock). Interstellar *gas* exists in a variety of chemical forms, i.e. molecular, atomic and ionized. Overall, it can be in one of five thermal phases (see also table 2.1). We start out with providing a brief overview of these phases, after which we zoom in on Molecular Clouds – the birth sites of stars and planets. Roughly in order from coolest to hottest and explicitly referencing the chemical form of the dominant element, hydrogen, we distinguish:

1. Very cold and dense Molecular Clouds (MCs): $H_2$
   Very cold ($T \sim 10$ K) and dense (particle number densities $n > 300\,\mathrm{cm}^{-3}$) molecular gas, distributed in *molecular coulds*. In our galaxy Molecular Clouds comprise $\sim$20 percent of the baryonic mass of the ISM, but occupy only $\sim$0.01 percent of its volume. It is on account of the (relatively speaking) high densities in these clouds that molecules can actually form, hence the name. More specifically, it is firstly because the processes forming molecules in interstellar space – primarily two-body gas-phase reactions and catalysis on the surface of dust grains – proceed faster at higher density, and secondly because dust effectively shields the interior of a dense cloud from the ultraviolet radiation which destroys molecules. These clouds are often 'dark' – with visual extinction $A_V \gtrsim 3$ mag through their central regions. The main tracers are millimeter-wavelength molecular emission lines (for instance CO). Dust grains in dark clouds are often coated with mantles of $H_2O$ and other molecular ices. Most molecular clouds are gravitationally bound. The densest regions are likely unstable and sites of new star formation. Note that the gas pressures in the densest regions would qualify as ultrahigh vacuum in a terrestrial laboratorium.

2. Cold Neutral Medium (CNM): H I, traced through spectral line absorption
   Cold ($T \sim 100$ K) gas is distributed in clouds, sheets and filaments of densities $n \sim 30\,\mathrm{cm}^{-3}$ occupying $\sim$1–4 percent of the ISM. The material is mainly neutral and atomic

(though species with low ionization potentials may be ionized). The densities are too low for molecules to form abundantly. These clouds are typically detected through the absorption of background stellar light in optical (e.g. Ca II, Na I and K I) and near ultraviolet (e.g. Mg II) ground-level resonance lines.

When densities and column densities are sufficiently large, $H_2$ may form and survive due to $H_2$ self-shielding. We refer to such clouds as diffuse molecular clouds. They are similar to the CNM, though temperatures may be somewhat lower ($\sim 50$ K) and densities somewhat higher ($\sim 100$ cm$^{-3}$). We do not rank them as a separate phase (but some do). They can be observed with the same diagnostics as the CNM and in addition by CO 2.6-mm emission.

3. Warm Neutral Medium (WNM): H I, traced through spectral line emission
   Warm neutral atomic hydrogen occupies $\sim$20–40 percent of the volume of the ISM and is located mainly in the boundary layers between molecular clouds and H II regions. It has characteristic temperatures of 5000–8000 K and densities of $\sim$0.5 cm$^{-3}$, and is traced by the H I emission line at 21 cm. These boundary layers are characterized by the photo-dissociation of the $H_2$ molecule and are therefore termed photodissociation regions. Some older literature refers to the Warm Neutral Medium as 'warm inter-cloud medium'.

4. Warm Ionized Medium: H II, traced through spectral line emission
   Diffuse gas with temperatures of 6000-10000 K, and densities $\sim$0.3 cm$^{-3}$ occupying about 15 percent of the volume of the ISM. While primarily photo-ionized (it requires about 1/6$^{\text{th}}$ of all the ionizing photons emitted by the Galaxy's O and B stars), there is some evidence of shock or collisional ionization high above the plane of the Galaxy. It is traced by low-surface brightness H$\alpha$ $\lambda$6563 Å emission.

   Nearly 90 percent of the H II in the Galaxy resides in the Warm Ionized Medium, with the remaining 10% in the bright high-density (compact, or at least localized) H II regions that occupy a tiny fraction of the ISM and that arise from the ultraviolet radiation of nearby, hot massive stars. The extended and diffuse Warm Ionized Medium (sometimes referred to as diffuse H II) is ionized by the background ultraviolet radiation field, also predominantly originating from hot massive stars. The Orion Nebula, measuring a few pc across, is an example of a bright H II region. Its lifetime is essentially that of the ionizing stars, so 3–10 Myr.

   In addition to the H II regions, photo-ionized gas is also found in *planetary nebulae* – these are created when rapid mass loss during the late stages of evolution of stars with initial mass $0.8 M_\odot < M < 8 M_\odot$ exposes the hot stellar core, creating a luminous planetary nebula. Individual planetary nebulae fade away on a $\sim 10^4$ yr time scale.

5. Coronal or Hot Ionized Medium (HIM): X-ray and O IV-VI absorption and emission
   Hot, low-density gas heated by shock waves from supernovae, with temperatures exceeding $> 3 \times 10^6$ K and very low densities of $< 0.004$ cm$^{-3}$, occupies $\sim$50 percent of the ISM. The vertical scale height of this gas is $\sim$3 kpc, so it is sometimes referred to in the literature as the hot 'corona' of the galaxy. This hot gas is often buoyant and

| Phase | $n_{\text{tot}}$ [cm$^{-3}$] | $T$ [K] | $M$ [$10^9 \, M_\odot$] | $f$ |
|---|---|---|---|---|
| molecular | $> 300$ | 10 | 2.0 | 0.01 |
| cold neutral | 50 | 80 | 3.0 | 0.04 |
| warm neutral | 0.5 | 8 000 | 4.0 | 0.30 |
| warm ionized | 0.3 | 8 000 | 1.0 | 0.15 |
| hot ionized | 0.003 | $5 \times 10^5$ | – | 0.50 |

Table 2.1: *Phases of the Interstellar Medium. From: Stahler & Palla, The Formation of Stars, with small adaptations. $f$ is the volume filling factor. The values are indicative only. Masses are given for our Milky Way galaxy.*

appears as bubbles, with characteristic dimensions of $\sim$20 pc, and fountains high above and below the disk. The coronal gas volumes may be connected to other coronal gas volumes. The gas is collisionally ionized and cools on $\sim$Myr timescales. Its primary tracers are absorption lines seen towards hot stars in the far-ultraviolet (e.g. O IV, N V, and C IV) in gas with $T \sim 10^5$ K, and diffuse soft X-ray emission from gas hotter than $10^6$ K. In the hottest coronal gas O VI is present.

Mixed in with all but the very hottest phases of the ISM is *interstellar dust*. Dust grains constitute only about 1 percent of the total mass of the ISM (the gas being responsible for the other 99 percent), but are an important source of interstellar extinction and gas-phase element depletion and are the sites of interstellar chemistry. Grains constitute material in the solid-phase, rather than in the gas-phase. Dust grains in the ISM range in size from a few microns down to macro-molecular scales (clusters of 50–100 atoms or less). While we will spend some time talking about gas-phase processes, it is important to keep dust in mind, since it plays a role often disproportionate to its share of the mass in the ISM.

Though molecules occur only in the densest regions of the interstellar medium, dust is wide-spread. Also lines of sight without intervening molecular clouds show appreciable extinction of starlight. Clearly, the dust cannot have been formed in situ in the low-density interstellar medium. The consensus is that dust originates in the outflows of stars during late stages of evolution. In these stages the stars often have cool surfaces and the stellar-wind speeds are low, presenting the best conditions for dust to form. Molecules formed in the outflows of Asymptotic Giant Branch (AGB) stars have a harder time than dust to survive in the interstellar medium. The diffuse ultraviolet radiation from hot stars in the galaxy and that of the emerging hot stellar core during the planetary-nebula phase (that over time will develop into a white dwarf) will quite efficiently photo-dissociate molecules deposited in the interstellar medium. Also dust is processed by high-energy radiation in the ISM, but it manages to survive in some form.

Save for the Hot Ionized Medium and some colder 'high-velocity clouds', the gas and dust clouds are confined to the galactic plane, with a scale height of the order of only a few tens of parsec. Interestingly, in the vicinity of the sun one observes that the main regions where fairly massive stars have been formed, i.e. recent star formation regions, define a warped structure above and under the plane referred to as *Gould's Belt*.

## 2.2 Observational overview of stellar nurseries: Molecular Clouds

Star formation occurs at small and large scales. We briefly discuss a number of well known star formation sites, from small to large. Some of the key properties of these sites are listed in table 2.2.

### 2.2.1 Taurus-Auriga

An **emission nebula** is a cloud of ionized gas. The most common source for ionizations are high energy photons emitted from a nearby star. These can be among some of the most colorful objects in the sky.

**Reflection Nebulae** are clouds of dust which are simply reflecting the light of a nearby star or stars. The nearby star or stars are not hot enough to cause ionization in the gas of the nebula like in emission nebulae but are bright enough to give sufficient scattering to make the dust visible.

Besides the luminous bright nebulae, there are **dark or absorption nebulae** which can be seen because they obscure, or absorb, the light coming from stars or bright nebulae behind them. A famous example is the **Horsehead Nebula.**

Figure 2.1: *Observations-inspired nomenclature used to describe nebulae.*

Taurus-Auriga is one of the closest moderately large star-forming regions. It is at a distance of about $140\,\mathrm{pc}$ and contains some $10^4\,M_\odot$ of molecular hydrogen. The region as a whole measures some $30\,\mathrm{pc}$ in linear extent. To get an impression of the apparent size in the sky, look at the well known constellation of Orion to the lower-left in figure 2.2 which is about half this size. Due to its closeness and its occurrence in the northern hemisphere, Taurus-Auriga has become a kind of archetypical region for the study of (low-mass) star formation. There are no high-mass stars in Taurus-Auriga, hence no H II-regions or emission nebulae (see figure 2.1); the highest-mass star in the region is not much more massive than the sun. With respect to higher-mass Star Formation Regions (SFRs) it appears that low-mass SFRs are less dense in existing low-mass stars. The low stellar density combined with its proximity makes Taurus-Auriga one of the best – if not the best – place to study star formation and young stars in relative isolation. The proto-typical pre-main-sequence star T Tauri is a member of the Taurus-Auriga Complex.

The Taurus-Auriga region contains a number of dark clouds, such as that of the dense core TMC 1 (Taurus Molecular Cloud number 1; see figure 2.3) that has been closely studied because of its wealth of interstellar molecules. These clouds appear as black spots in optical images because the dust extinction in such clouds is so large that the light from stars in the
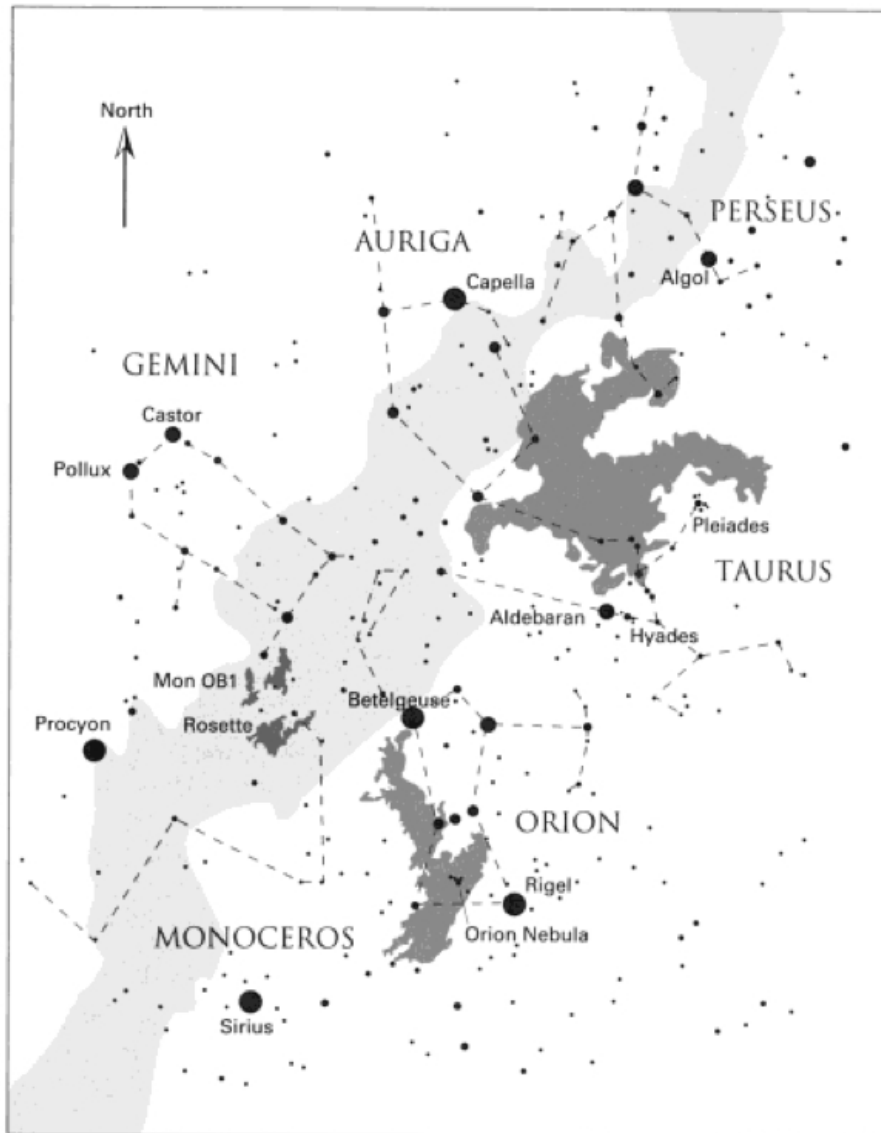
Figure 2.2: *A portion of the Northern sky. The Milky Way is depicted in light grey, while the darker patches indicate (giant) molecular clouds. The Taurus-Auriga MC is at a distance of 140 pc and measures about 30 pc in linear extent. The Orion GMC is some 450 pc distance and has a linear size of about 120 pc. Also shown, according to their relative brightness, are the more prominent stars, along with principle constellations. From: Stahler & Palla, The Formation of Stars.*
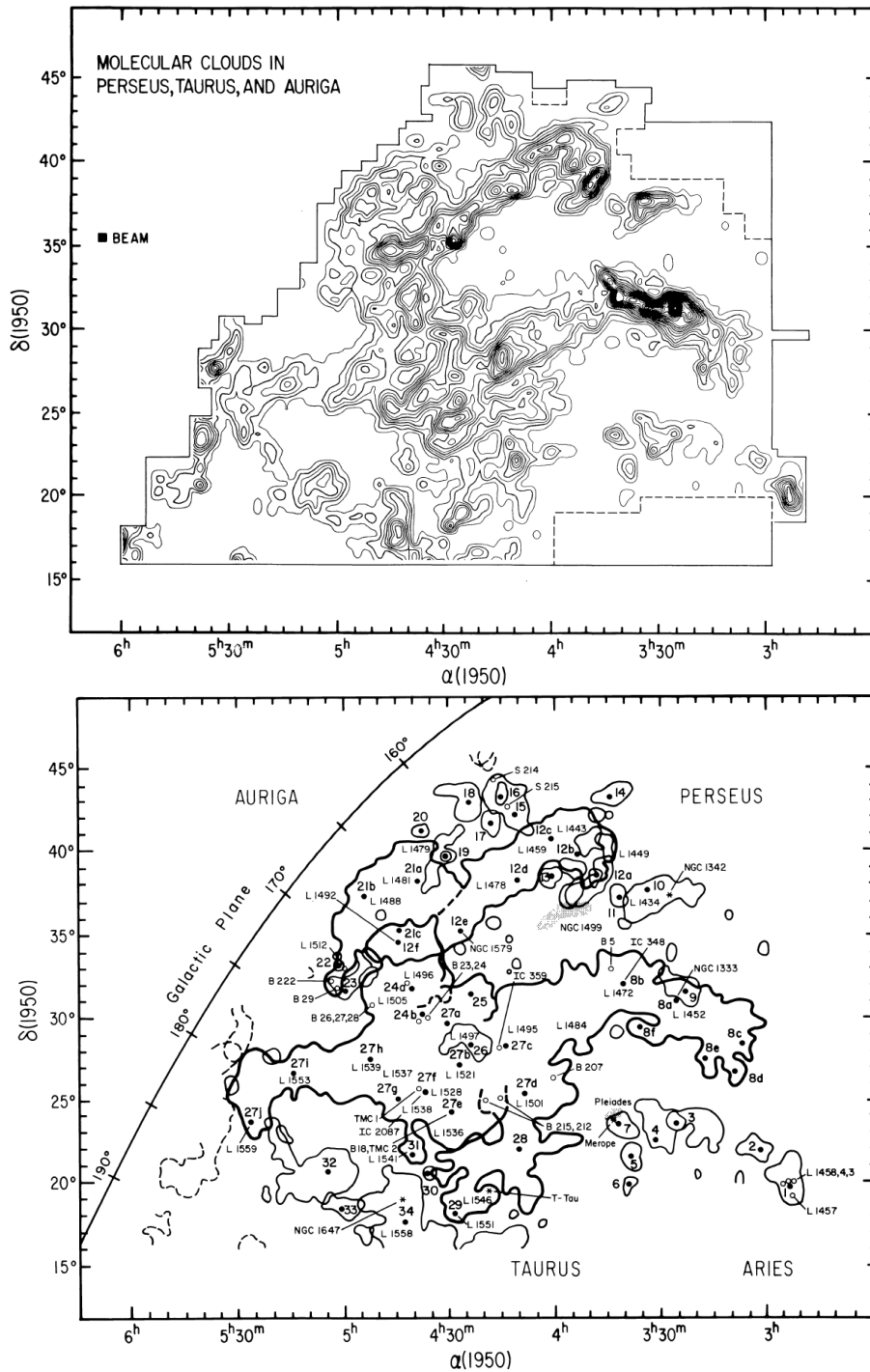
Figure 2.3: *An overview structure in Taurus-Auriga. Top: velocity-integrated intensity of CO emission. The lowest contour is $0.5\,K\,km\,s^{-1}$, and the separation between contours is $1.5\,K\,km\,s^{-1}$. The border of the surveyed region is indicated by the outer, solid line. Bottom: designation of clouds and CO peaks (full circles; large labels) and locations of various objects from the NGC, IC, Lynds, and Barnard catalogs (small labels). Cloud outlines correspond to the $0.5\,K\,km\,s^{-1}$ level of the velocity-integrated intensity. From: Ungerechts & Thaddeus 1987, ApJS 63, 645.*

| Cloud Type | $A_V$ | $n$ | $D$ | $T$ | $M$ | Examples |
| --- | --- | --- | --- | --- | --- | --- |
| | mag | cm$^{-3}$ | pc | K | $M_\odot$ | |
| Diffuse | 1 | 500 | 3 | 50 | 50 | $\zeta$ Ophiuchi |
| Giant Molecular Cloud | 2 | 100 | 50 | 15 | $10^5$ | Orion, 30 Doradus |
| Dark Clouds | | | | | | |
|     Complexes | 5 | 500 | 10 | 10 | $10^4$ | Taurus-Auriga |
|     Individual | 10 | $10^3$ | 2 | 10 | 30 | B1 |
| Dense Cores / Bok Globules | 10 | $10^4$ | 0.1 | 10 | 10 | TMC-1/B335, B68 |

Table 2.2: *Typical physical properties of Molecular Clouds. From: Stahler & Palla, The Formation of Stars.*

background is completely absorbed (and because Taurus-Auriga is so nearby that no bright stars are in the foreground). Note that the elongated structure to the north, including L 1459 and L 1434, and the high-intensity area to the west containing Barnard 5 (B5) and NGC 1333, are not physically associated with the Taurus-Auriga system, but are more distant, massive clouds seen in projection.

## 2.2.2  $\rho$ Ophiuchi cloud

If observational astronomy had started in the southern hemisphere, the $\rho$ Ophiuchi cloud in the southern constellation Ophiuchus may have been the archetypical star-forming region. It is located at 150 pc, so about the same distance as the Taurus-Auriga cloud, but is somewhat less massive (containing $\sim 3 \times 10^3 \, M_\odot$).

Figure 2.4 shows the dark cloud Barnard 68 in the $\rho$ Ophiuchi cloud. This star forming region is so nearby that not a single bright star is situated between it and us. The reason that this cloud is dark is due to dust extinction. The extinction is particularly strong at visible light. The Very Large Telescope has revealed the presence of about 3 700 background Milky Way stars of which the light is blocked by the dust in B68, some 1 000 of which are visible at infrared wavelengths. The cloud's mass is about twice that of the sun, and it measures about half a light-year across. Barnard's 68's well-defined edges and other features show that it is on the verge of gravitational collapse within the next 100 000 years or so, and is on its way to becoming a star or multiple system.

## 2.2.3  Orion Complex and the Orion Nebula

The Orion Complex is situated south of Orion's Belt in the constellation of Orion (see the dark grey area in figure 2.2) and is the closest example of a Giant Molecular Cloud complex. With a linear size of about 120 pc and a distance of 450 pc, it covers about 12 degrees on

B, V, I                                                                    B, I, K
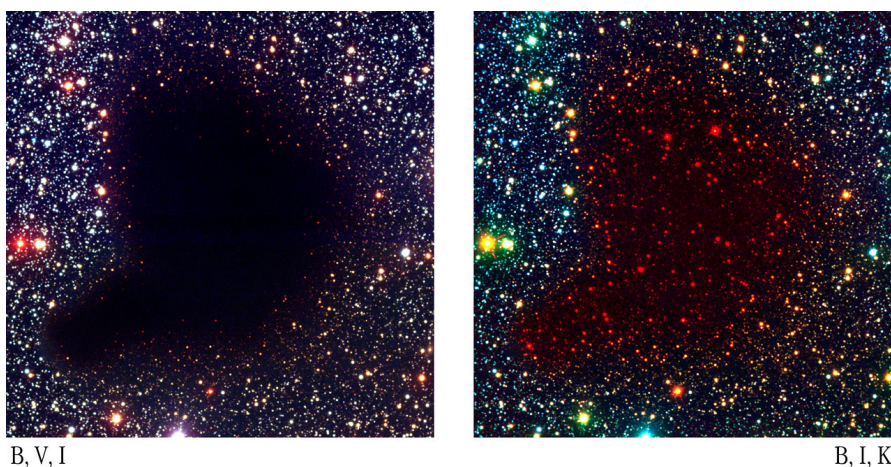
Figure 2.4: *A dark cloud that is* not *embedded within a larger complex is referred to as a* Bok Globule, *named in honor of the astronomer Bart Bok who first recognized, in the 1940s, their potential role in stellar birth. The figure shows a comparison of the central area of Bok Globule Barnard 68 (B68) in a color composite of visible and near-infrared on the left and a false-color composite based on a visible (here rendered as blue), a near-infrared (green) and an infrared (red) image on the right. See also figure 5.1.*

the sky. The total amount of mass in the Orion Complex is estimated at several $10^6 \, M_\odot$. The distinction between a Molecular Cloud and a Giant Molecular Cloud is poorly defined, however, a workable definition of a GMC can be that the cloud contains sufficient mass (more than roughly $10^5 \, M_\odot$) to also produce massive (i.e. $M \gtrsim 9 M_\odot$) stars. When young massive stars (of spectral type O and early-B) are present, the ambient hydrogen gas is ionized and one or more H II regions are created. This too then distinguishes MCs from GMCs.
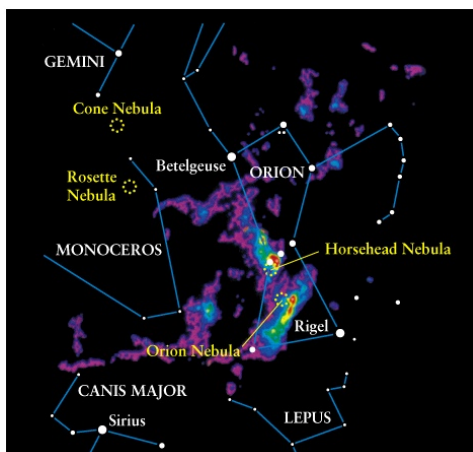


Figure 2.5: *The Orion Complex, outlined by the emission of CO at a wavelength of 2.6 mm.*

Several sub-units can be distinguished in the Orion Complex; indeed, there is no reason why all stars in the same cloud would be co-eval, and for massive stars such as formed in Orion the total lifetime (of $\sim 1-2\times 10^7$ yr) is not necessarily much longer than the total duration of the star-forming activity. Hence the occurrence in a same complex of well evolved stars such as Betelgeuze, crowded massive-star clusters such as the Trapezium Cluster in the densest part of the Orion Nebula, and deeply embedded sources. A famous object among the latter is the 'BN' (for Becklin-Neugebauer) object in the 'Orion-KL' (for Kleinmann-Low) nebula, a powerful infrared source which produces a massive outflow, affecting the environment.
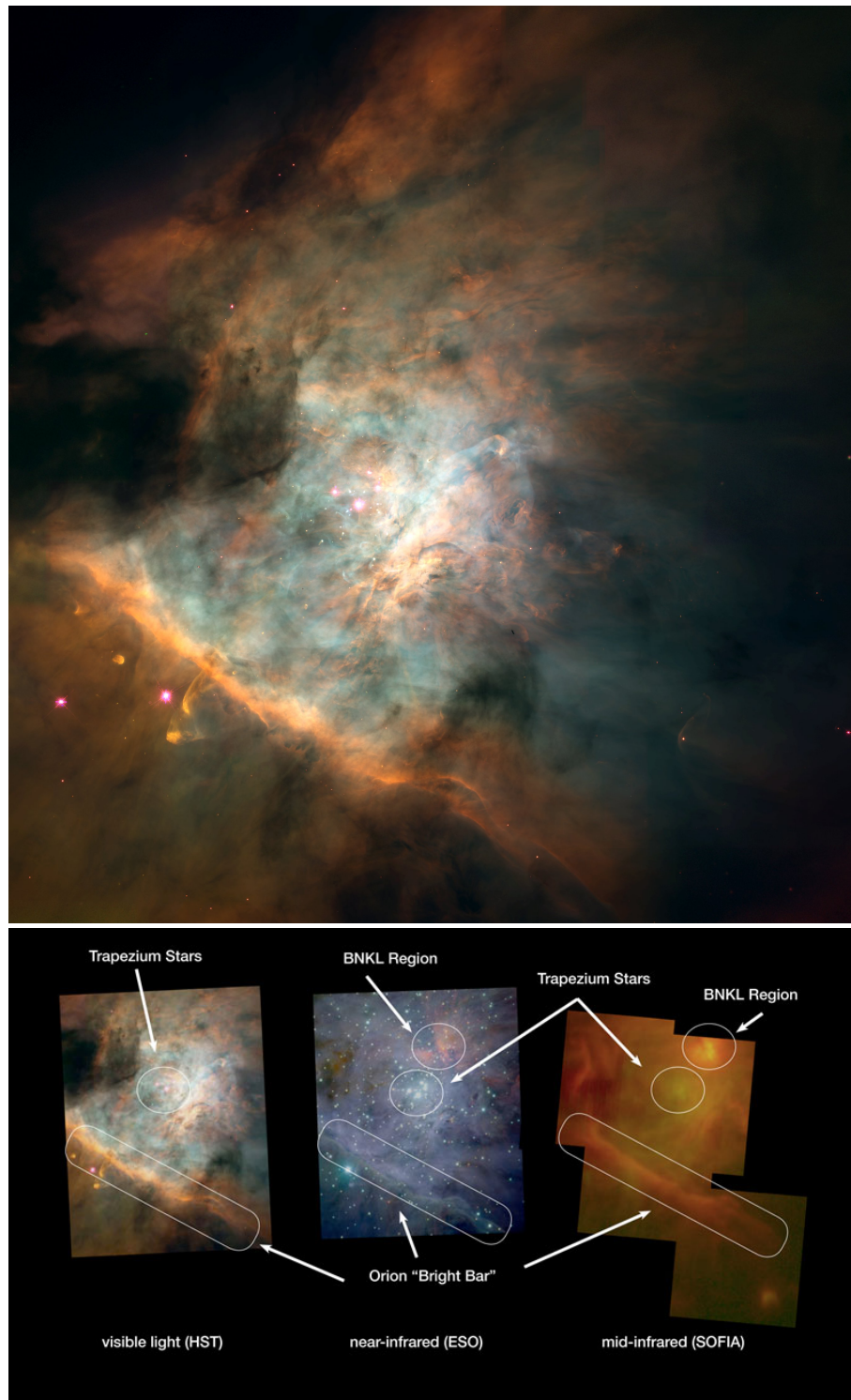
Figure 2.6: *The Orion Nebula (Messier 42) is one of the brightest nebula and is visible to the naked eye in the night sky. It is the closest site of massive star formation to earth. In the bottom panels different structures in the nebula are highlighted: the Trapezium stars, the Orion 'Bright Bar', and the Becklin-Neugebauer / Kleinmann-Low (BNKL) region.*

Arguably the most famous sub-unit in the Orion Complex is the Orion Nebula (rivaled perhaps by the Horsehead Nebula), also known as Messier 42, M42 or NGC 1976. The Orion Nebula measures 24 lightyear across and contains about 2 000 solar masses of material. It contains a very young open cluster known as the Trapezium. The five brightest stars – all of which are multiple – have masses of 15 to 30 solar masses. They are within a projected diameter of 1.5 lightyears of each other and are responsible for much of the illumination of the surrounding nebula. The Trapezium is a sub-component of the larger Orion Nebula Cluster, a grouping of about 2 000 stars within a diameter of 20 lightyears.

The Orion Bar lies at the boundary between the relatively sparse (gaseous) region in the center of the images shown in figure 2.6 and a cold dense cloud of gas and dust at the lower left. The Orion Bar is an *ionization front* or *photo-dissociation region*, meaning it is the outer edge of the bubble of gas that is being heated by ultraviolet radiation from the Trapezium stars, expanding away from them and 'eating into' the cold dense cloud. The visible light image shows the Orion Bar as a mostly opaque cloud illuminated from the side by the Trapezium stars. The near-infrared image sees partly into the Orion Bar, and shows excited gas within the Orion Bar. The mid-infrared view reveals emission from dust in the Orion Bar, including emission from Polycyclic Aromatic Hydrocarbons (PAHs).

Located in the upper right of these images, the BN/KL region contains a number of deeply embedded protostars of which the Becklin-Neugebauer object is the brightest (though not at wavelengths $\geq 31.5\,\mu$m; De Buizer et al. 2012). The BN object will eventually become a B-type stars, signifying it as the nearest location of ongoing high-mass star formation. The adjacent Kleinmann-Low nebula is a cooler cloud containing multiple sources and will eventually form multiple stars that – given the compact nature of the region – may perhaps suffer protostellar collisions. No trace can be seen of BN/KL in the visible-light image.

### 2.2.4  30 Doradus

The Orion GMC is one of the thousand or so giant molecular clouds in our Galaxy. The most massive galactic molecular cloud is Sgr B2, at some 200 pc from the galactic center. Still closer to the Milky Way center are the Arches and Quintuplet clusters, which contain many O stars. Somewhat closer to us, in the Carina spiral arm, is NGC 3603, an H II region containing a very massive cluster, with about 100 times the ionizing luminosity of the Trapezium Cluster. The largest star forming region in the local universe is the 30 Doradus Nebula in the LMC; put at the distance of Orion it would fill 10 percent of the sky! Some useful nomenclature regarding 30 Doradus is provided in table 2.3. An overview of the LMC is given in figure 2.7, which shows that 30 Doradus is the most prominent feature in its host galaxy.

Hubble's 22nd Anniversary Image, figure 2.8[1], captures the extreme complexity of the 30 Doradus star forming region, also known as the Tarantula Nebula. It measures hundreds of light years across and harbors several young clusters of various ages, active star forming regions containing dark clouds, proto-stars and pre-main sequence stars (objects of the latter two

---

[1]HST-GO 12499 (PI: D. Lennon). See http://hubblesite.org/newscenter/archive/releases/2012/01/

|                          | Linear scale | Angular scale |          |          |
|                          | [pc]         | at 52.5 kpc   | at 1 Mpc | at 10 Mpc |
|--------------------------|--------------|---------------|----------|-----------|
| LMC                      | 5 000        | 5°            | 15'      | 1.5'      |
| 30 Doradus Region        | 1 000        | 1°            | 3'       | 20"       |
| 30 Doradus Nebula        | 200          | 15'           | 50"      | 5"        |
| 30 Doradus Cluster       | 40           | 3'            | 10"      | 1"        |
| Radcliffe 136 (R136)     | 2.5          | 10"           | 0.5"     | 0.05"     |
| Radcliffe 136a (R136a)   | 0.25         | 1"            | 0.05"    | 0.005"    |

Table 2.3: *The 30 Doradus Nebula is also known as the Tarantula Nebula or NGC 2070. To place this star forming region in perspective several scale (both larger and smaller) are listed, both in linear measure, and in angular measure for the LMC and greater distances. The 30 Doradus Region is usually taken to encompass Henize. From: Walborn (1991), IAUS 148, 145.*

classes are often termed Young Stellar Objects or YSOs). We refrain from a discussion of objects that have not yet reached the main sequence (which has been covered above) and focus on the rich population of short lived massive stars

Star formation in this region has been ongoing for several tens of millions of years reaching a peak some 1–2 Myr ago with the formation of the cluster Radcliffe 136, at the core of 30 Doradus and currently the dominating ionizing source of the region. The 30 Doradus Region contains hundreds of thousands of stars, of which several thousand are massive stars. The very center of R136 contains a number of stars of spectral type WNh, identified as extremely massive main-sequence stars with stellar winds so dense that their optical spectra are dominated by emission lines, similar to that of evolved (i.e. post main-sequence) Wolf-Rayet stars. These are the most massive stars known in the universe, with initial masses perhaps as high as 165–320 $M_\odot$ (Crowther et al. 2010). It is a mystery how such stars may have formed.

Within a radius of 4.7 pc (or 19.4") of the dense R136 cluster center the mass in stars is estimated to be $\sim 10^5$ $M_\odot$ if the initial mass function is extrapolated down to 0.1 $M_\odot$. It can probably be considered a small young globular cluster. A total mass of $\sim 10^5$ $M_\odot$ qualifies R136 as a super star cluster, albeit a rather modest one in a cosmological context as more distant star-bursing dwarf galaxies are known to host super star clusters ten times as massive. Extreme star-bursting systems, such as the interacting Antennae Galaxies, may contain even more massive clusters, reaching $\sim 10^7$ $M_\odot$.

Several associations can be seen in figure 2.8. Some 3' to the north of R136 lies the association Hodge 301, hosting A- and M-type supergiants. The association NGC 2060 is located some 6.5" to the southwest of R136. In addition to O and B stars, it hosts a number of luminous red supergiants and the supernova remnant B0538-691. Finally, the association SL 639 is located 7.5" to the southeast. Its brightest members are of spectral type B.
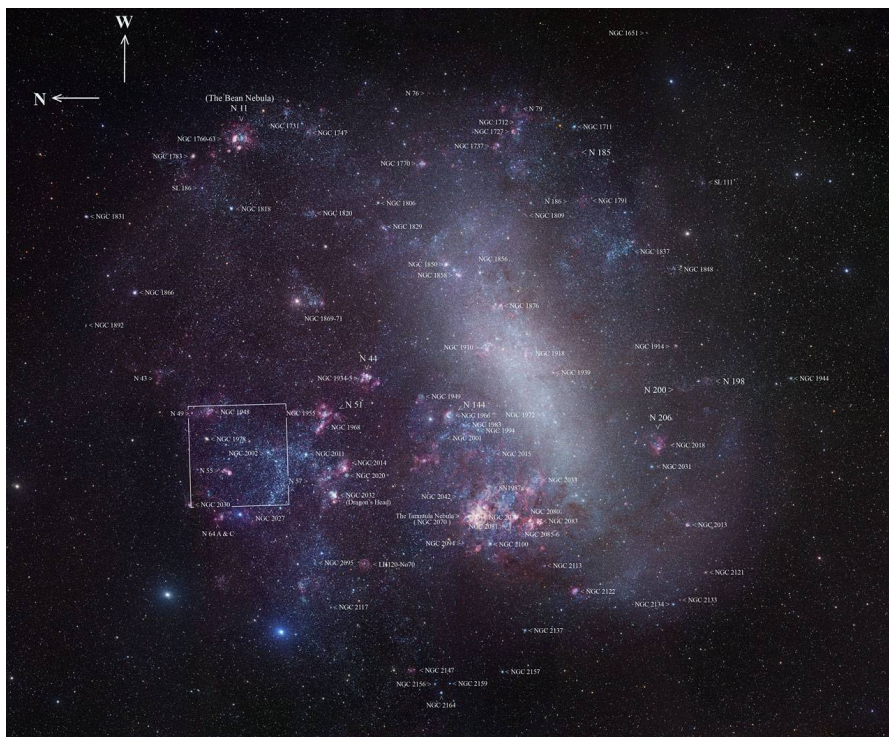
Figure 2.7: *An overview of the Large Magellanic Cloud. The image covers a region of sky of 645 x 550 arc-minutes (recall the diameter of the moon being 31 arc-minutes). Dozens of clusters of young stars can be seen as well as traces of gas clouds. In addition to the Tarantula Nebula (NGC 2070) other star forming sites can be seen that are also part of the ∼1 kpc sized 30 Doradus Region. The image clearly shows that the LMC is remarkably active in star formation. From: WPI on the MPG/ESO 2.2m telescope at La Silla Observatory.*

### 2.2.5 M51

Figure 2.9 shows the Whirlpool grand-design spiral galaxy (M51, also known as NGC 5194) at a distance of 7 Mpc. The galaxy is known to be undergoing an interaction with its companion NGC 5195, and shows a system of star formation, largely confined to two spiral arms. The tilt angle of M51 is 20°, so we see it almost face on. Dynamical simulations of the interaction suggest two recent disk passages at 400-500 Myr and 50-100 Myr ago. M51 shows similarities with our own galaxy, thus illustrating that also our galaxy may contain very many giant molecular clouds such as Orion or 30 Doradus. The bottom panel of the figure provides ages, masses and location of 120 star clusters. Such parameters could not be determined for most star clusters, though it was established that M51 hosts at least 900 such young massive clusters – all star clusters with ages of ∼ 10 Myr or less are associated with GMCs.

This is where we stop our overview of star forming regions. At much larger distances it becomes hard, or even impossible, to spatially resolve galaxies. Still, the star formation activity

Figure 2.8: *Compilation of images of the Giant Molecular Clould complex 30 Doradus in the Large Magellanic Cloud taken with Hubble Space Telescope to celebrate the 22nd anniversary of the space observatory. The top figure is the actual image; the bottom figure is an annotated black-and-white version explaining various structures that typically are present at such a large star-formation site.*

of such unresolved galaxies can be studied by means of the integrated light of these systems. Such studies focus on the fraction of this total light that is produced by young massive stars, which fortunately are very bright (see section 1.2) and invariably associated with star formation. On the biggest – cosmological – scale one may thus derive the cosmic history of star formation. This history is shown in figure 1.4 and briefly discussed in section 1.3.
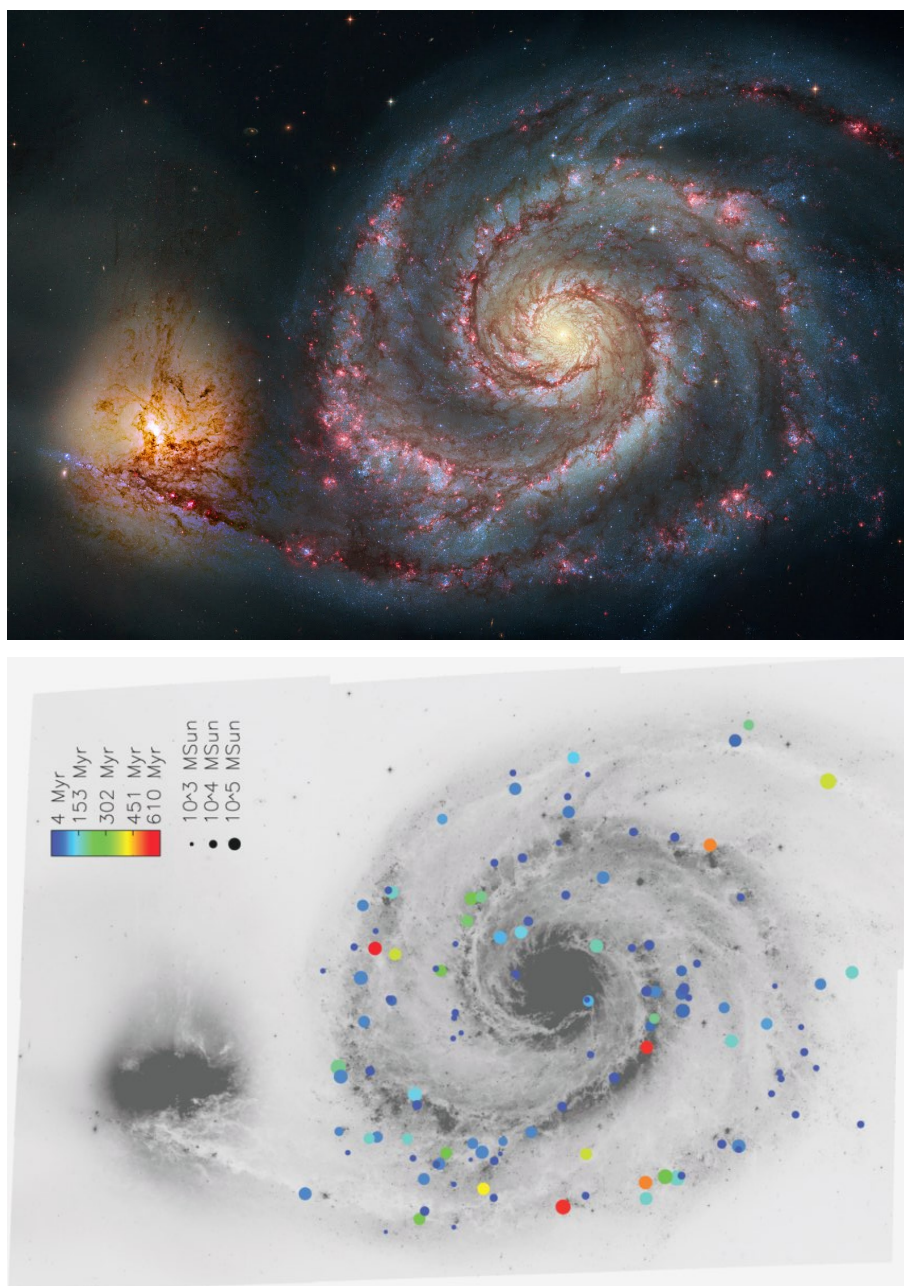
Figure 2.9: *A map of M51 indicating the age, mass and location of 120 associations deemed to single-aged. M51 hosts at least 900 associations, but in the large majority of these it is difficult to identify the main sequence turnoff from photometry, potentially because of multiple massive associations contributing to the integrated light of the identified sources and/or the limited sensitivity of the photometric images. Note that this galaxy, which shows similarities with our Milky Way galaxy, indeed contains at least $\sim 1000$ giant molecular clouds. From: Kaleida & Scowen 2009, AJ 140, 379.*

**Exercise 2.1**

Table 2.1 summarizes key properties of the different phases of the ISM. Use eq. 3.2 to estimate whether these different ISM phases are in pressure equilibrium with one another. Acknowledging that the numbers are rough estimates, assume that pressure equilibrium hold when the pressure difference is not larger than a factor of three.

**Exercise 2.2**

If you were asked to order R136, Hodge 301, NGC 2060 and SL 639 in terms of cluster/association age, from young to older, what order would you give? Motivate your answer.

# 3

# On the verge of collapse

To understand the process of star formation, we must first understand the physical processes which take place in molecular clouds. In this chapter we will look at the stability of molecular clouds and begin to assess what might be needed to make such clouds collapse under gravity and form stars.

## 3.1 Fluid dynamics and the equation of state

### 3.1.1 Simple equation of state

The behavior of a gas or solid under different physical conditions is described by the equation of state (or EOS), expressing how the pressure depends on density and temperature

$$p = f(\rho, T). \tag{3.1}$$

The simplest assumption for the EOS of the gas in a molecular cloud is that it behaves as an ideal gas, i.e.

$$p = n\,k\,T = \frac{\rho\,k\,T}{\mu m_\mathrm{H}} = \frac{\rho\,\mathcal{R}\,T}{\mu}, \tag{3.2}$$

where $p$ is the gas pressure in dyn/cm$^2$, $n$ is the particle density in cm$^{-3}$, $\rho$ is the density in gr/cm$^3$, $T$ is the temperature in K, and $\mu$ is the mean molecular weight of the gas in units of $m_\mathrm{H}$, which would be two if the cloud were entirely composed of molecular hydrogen. It holds that $\rho = \mu\,m_\mathrm{H}\,n$. The constants have their usual meaning; $\mathcal{R} = k/m_\mathrm{H}$ is the gas constant. It is sometimes convenient when treating astrophysical problems to assume that the gas temperature and chemical composition remain constant. The former is known as the isothermal assumption, under which equation 3.2 becomes

$$\frac{p}{\rho} = \frac{k\,T}{\mu m_\mathrm{H}} = \frac{\mathcal{R}\,T}{\mu} = a^2 = \mathrm{constant}, \tag{3.3}$$

where $a$ is the isothermal sound speed in the gas at temperature $T$. In general, the sound speed is given by $a^2 = dp/d\rho$.

The next simplest assumption we can adopt is the adiabatic equation of state

$$p = K\rho^\gamma, \tag{3.4}$$

where $K$ is constant and $\gamma$ is the ratio of specific heats at constant pressure, $c_p$, and volume, $c_v$, respectively: $\gamma = c_p/c_v$. An adiabatic process is any process occurring without gain or loss of heat within a system, i.e. during the process the system is thermodynamically isolated. For an adiabatic process, $a^2 = \gamma p/\rho$.

It is fairly straightforward to derive the adiabatic index for a gas that is pure in the sense that it consists (or is dominated by) one type of particle. In that case it can be related to the degrees of freedom $f$ of the molecule or atom/ion,

$$\gamma = 1 + \frac{2}{f}. \tag{3.5}$$

For a monoatomic gas, for instance, with three (translational) degrees of freedom, $\gamma = 5/3$. For a diatomic gas, two rotational degrees of freedom may be added[1], yielding $\gamma = 7/5$. At high temperatures a vibrational degree of freedom may be added to diatomic molecules as well. For gasses that consist of different components, for instance a gas that is partly molecular and partly atomic and or ionic, it is more complex to determine the value of $\gamma$.

### 3.1.2  Conservation of mass

The material in an astrophysical medium may not be static (i.e., $\partial/\partial t = 0$ and $v = 0$) or stationary (i.e., $\partial/\partial t = 0$), and for a general description we must take this into account. To do so, we use the basic equations of aero/fluid mechanics. In general, there are two methods commonly used to model gas/fluid flow. One method is to use a fixed set of coordinates in space and calculate the parameters of the gas as it flows through the coordinate frame. This is known as the Eulerian method. An alternative is to choose a set of coordinates fixed to a particle of the gas, moving with that particle, and to calculate the varying parameters in that coordinate frame (referred to as co-moving coordinates). This is known as the Lagrangian method.

Assume an arbitrary volume $V$ contained by a closed surface $S$. $V$ and $S$ are fixed in space, so we are here adopting the Eulerian viewpoint. The mass flowing out of $V$ per unit time through the element of area $\mathbf{dS}$ of the surface is given by

$$\rho\mathbf{v} \cdot \mathbf{dS}, \tag{3.6}$$

where $\mathbf{v}$ is the macroscopic (i.e. systemic) velocity in cm sec$^{-1}$ and $\rho$ the total mass density in gr cm$^{-3}$, and so the net rate at which mass flows out of $V$ through $S$ is given by

$$\oint_S \rho\mathbf{v} \cdot \mathbf{dS} = \int_V \nabla \cdot (\rho\mathbf{v}) \, dV, \tag{3.7}$$

---

[1]These rotation axes are both perpendicular to the bond axis, i.e. the line connecting the atoms.

where we have obtained the right-hand side by invoking Gauss's divergence theorem. The rate at which the mass in $V$ decreases is given by

$$-\frac{\partial}{\partial t}\left(\int_V \rho dV\right) = \int_V \left(-\frac{\partial \rho}{\partial t}\right) dV, \tag{3.8}$$

where we can take $\partial/\partial t$ inside the integral because $V$ is fixed in space. Obviously, the rate at which the mass in $V$ decreases, i.e. eq. (3.8), must equal the rate at which mass flows out of $V$, across $S$, so

$$\int_V \left[\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v})\right] dV = 0, \tag{3.9}$$

and since the volume $V$ is arbitrary, it follows that

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \tag{3.10}$$

everywhere. This is known as the *mass continuity equation*. We can expand this equation and rewrite it as

$$\frac{\partial \rho}{\partial t} + \rho \nabla \cdot \mathbf{v} + \mathbf{v} \cdot \nabla \rho = \frac{\partial \rho}{\partial t} + \rho \frac{\partial v^i}{\partial x^i} + v^i \frac{\partial \rho}{\partial x^i} = 0 \tag{3.11}$$

where the double occurrence of index $i$ in both terms in the second right-hand-side implies a summation over the components of the coordinate system.

We stress that $\partial \rho / \partial t$ is the Eulerian time derivative of the density, i.e. the rate of change of density at a fixed point in space. If we want the Lagrangian time derivative of the density, i.e. the rate of change of density moving with the fluid, we must include the contribution due to the displacement, $\mathbf{dr} = \mathbf{v}dt$, which occurs during the time interval $dt$. The net density change is

$$d\rho = \frac{\partial \rho}{\partial t}\, dt + \mathbf{dr} \cdot \nabla \rho = \frac{\partial \rho}{\partial t}\, dt + dx^i \frac{\partial \rho}{\partial x^i} \tag{3.12}$$

and hence the Lagrangian time derivative of the density is

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \frac{\partial \rho}{\partial t} + \mathbf{v} \cdot \nabla \rho = -\rho \left(\nabla \cdot \mathbf{v}\right), \tag{3.13}$$

where the final expression is obtained by substituting from equation (3.11)[2]. The Lagrangian time derivative is called variously the 'co-moving time derivative', 'fluid-frame derivative', 'total derivative', or 'material derivative'. Though here we use the symbol $\mathrm{d}/\mathrm{d}t$ for the Lagrangian derivative, it is more common to use the notation $\mathrm{D}/\mathrm{D}t$, to distinguish it from the ordinary derivative of a function of one variable in the sense it is usually used in mathematics. We do not do that. Here $\mathrm{d}/\mathrm{d}t$ is the rate of change of some characteristic (in the above case, density) of a particular element of fluid.

---

[2]If the density of a control volume which moves with the flow remains constant, i.e. the fluid is neither compressed or expanded, the flow is said to be incompressible. It is equivalent to a vanishing divergence of the flow velocity, i.e. $\nabla \cdot \mathbf{v} = 0$.

CARTESIAN COORDINATES $(\mathbf{x}, \mathbf{y}, \mathbf{z})$

A vector field $\quad\quad\quad\quad\quad \mathbf{A} = (A_x, A_y, A_z) = A_x\mathbf{x} + A_y\mathbf{y} + A_z\mathbf{z}$

Gradient $\quad\quad\quad\quad\quad\quad \nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right) = \frac{\partial f}{\partial x}\mathbf{x} + \frac{\partial f}{\partial y}\mathbf{y} + \frac{\partial f}{\partial z}\mathbf{z}$

Divergence $\quad\quad\quad\quad\quad \nabla \cdot \mathbf{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z}$

Differential displacement $\quad d\mathbf{s} = dx\,\mathbf{x} + dy\,\mathbf{y} + dz\,\mathbf{z}$

Differential volume $\quad\quad\quad dV = dx\,dy\,dz$

SPHERICAL COORDINATES $(\mathbf{r}, \boldsymbol{\theta}, \boldsymbol{\phi})$

A vector field $\quad\quad\quad\quad\quad \mathbf{A} = (A_r, A_\theta, A_\phi) = A_r\mathbf{r} + A_\theta\boldsymbol{\theta} + A_\phi\boldsymbol{\phi}$

Gradient $\quad\quad\quad\quad\quad\quad \nabla f = \left(\frac{\partial f}{\partial r}, \frac{\partial f}{\partial \theta}, \frac{\partial f}{\partial \phi}\right) = \frac{\partial f}{\partial r}\mathbf{r} + \frac{1}{r}\frac{\partial f}{\partial \theta}\boldsymbol{\theta} + \frac{1}{r\sin\theta}\frac{\partial f}{\partial \phi}\boldsymbol{\phi}$

Divergence $\quad\quad\quad\quad\quad \nabla \cdot \mathbf{A} = \frac{1}{r^2}\frac{\partial(r^2 A_r)}{\partial r} + \frac{1}{r\sin\theta}\frac{\partial(A_\theta \sin\theta)}{\partial \theta} + \frac{1}{r\sin\theta}\frac{\partial A_\phi}{\partial \phi}$

Differential displacement $\quad d\mathbf{s} = dr\,\mathbf{r} + rd\theta\,\boldsymbol{\theta} + r\sin\theta d\phi\,\boldsymbol{\phi}$

Differential volume $\quad\quad\quad dV = r^2\sin\theta\,dr\,d\theta\,d\phi$

## Mass-loss rate and mass-accretion rate

For a stationary one-dimensional radial spherical flow the mass continuity equation reduces to

$$\frac{1}{r^2}\frac{\partial(r^2\rho v)}{\partial r} = 0 \tag{3.14}$$

from which we find

$$4\pi r^2\rho(r)v(r) = \text{constant} \equiv \dot{M} \tag{3.15}$$

where $\dot{M}$ is the *mass-loss rate* through a spherical surface of radius $r$, and $v$ is the velocity in the radial direction. If the velocity $v$ is negative, i.e. directed towards the center of the sphere, $\dot{M}$ is to be interpreted as the *mass-accretion rate*. The cgs unit of mass loss or mass accretion is gr sec$^{-1}$, although it is custom to use the unit $M_\odot$ per year: $1\ M_\odot\text{yr}^{-1} = 6.303 \times 10^{25}$ gr sec$^{-1}$.

### 3.1.3 Conservation of momentum: the equation of motion under pressure

Consider again the volume of gas $V$. If the gas in this volume has pressure $p(\mathbf{r}, t)$, then the total force acting on the volume is the sum of the external pressure on the surface. This is given by the surface intergral

$$- \oint_S p \, dS. \tag{3.16}$$

Transforming this into a volume integral using Gauss's divergence theorem, the net pressure force exerted on the arbitrary volume $V$ is

$$- \oint_S p \, dS = - \int_V \nabla p \, dV, \tag{3.17}$$

and hence the net pressure force per unit volume is simply $-\nabla p$.

The equation of motion of this volume can be derived by equating the force per unit volume with the mass per unit volume multiplied by its acceleration. This is simply Newton's third law. The mass per unit volume is defined as the density $\rho$ and the acceleration is the time derivative of the velocity, i.e. $d\mathbf{v}/dt$. So we have

$$-\nabla p = \rho \frac{d\mathbf{v}}{dt}, \tag{3.18}$$

and hence

$$\frac{d\mathbf{v}}{dt} = -\frac{\nabla p}{\rho}. \tag{3.19}$$

Here $d\mathbf{v}/dt$ is the co-moving acceleration of the gas, so equation 3.19 is the Lagrangian formulation of the equation of motion. The Eulerian formulation is obtained by substituting for $d\mathbf{v}/dt$ using

$$\frac{d\mathbf{v}}{dt} = \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \, \mathbf{v}. \tag{3.20}$$

For the $j$-th component of velocity vector $\mathbf{v}$ this implies

$$\frac{dv_j}{dt} = \frac{\partial v_j}{\partial t} + v^i \frac{\partial v_j}{\partial x^i}. \tag{3.21}$$

### 3.1.4 Fluid motion in a gravitational potential field

If the gas is in a gravitational potential field $\Phi$ (with dimension $\mathrm{cm}^2 \, \mathrm{s}^{-2}$), where any gas element suffers an acceleration

$$\mathbf{g} = -\nabla \Phi, \tag{3.22}$$

an extra term must be included in eq. (3.19) to account for this. The force $\mathbf{f}_\mathrm{g}$ on unit volume due to a gravitational acceleration $\mathbf{g}$ is simply $\rho \, \mathbf{g}$, and so eq. (3.19) becomes

$$\frac{d\mathbf{v}}{dt} = \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \, \mathbf{v} = -\frac{\nabla p}{\rho} + \mathbf{g} \tag{3.23}$$

This is sometimes referred to as Euler's equation. Note the notation $(\mathbf{v} \cdot \nabla) \mathbf{v}$. This has by definition the components $v_i = \mathbf{v} \cdot (\nabla v_i)$, for $i = 1, 2, 3$, where $\mathbf{v} = (v_1, v_2, v_3)$. In spherical symmetry, with $\mathbf{v} = (v, 0, 0)$, this reduces to $v \, \partial v / \partial r$. The acceleration due to a mass $M$, which also implies spherical symmetry, is given by

$$\mathbf{g} = -\nabla \Phi = -\frac{\partial}{\partial r} \left( \frac{-GM}{r} \right) = -\frac{GM}{r^2}. \tag{3.24}$$

### 3.1.5  Fluid motion in a magnetic field

There is ample observational evidence that molecular clouds are permeated by magnetic fields. Electric currents in the fluid can give rise to a force acting on the gas, which then needs to be taken into account in the equation of motion. We begin with two of Maxwell's equations. Firstly we use Ampère's law

$$\nabla \times \mathbf{B} = \frac{4\pi}{c} \mathbf{J} - \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t}, \tag{3.25}$$

where $\mathbf{B}$ is the magnetic flux density, $\mathbf{J}$ is the current density, $\mathbf{E}$ is the electric field strength, and $c$ is the speed of light in a vacuum. Secondly we use Ohm's law, which for a fluid of conductivity $\sigma$ becomes

$$\mathbf{J} = \sigma \left[ \mathbf{E} + \frac{1}{c} \left( \mathbf{v} \times \mathbf{B} \right) \right]. \tag{3.26}$$

In this equation the first term on the right-hand side describes the conductivity of a static fluid and the second term incorporates the motion of that fluid at velocity $\mathbf{v}$. In the limit of infinite conductivity we have

$$\mathbf{E} + \frac{1}{c} \left( \mathbf{v} \times \mathbf{B} \right) = 0. \tag{3.27}$$

In a conducting fluid there are electric currents which can flow and give a force per unit volume $\mathbf{f}_{\mathrm{m}}$ given by

$$\mathbf{f}_{\mathrm{m}} = \frac{1}{c} \left( \mathbf{J} \times \mathbf{B} \right). \tag{3.28}$$

We can eliminate $\mathbf{J}$ from this force equation by using Ampère's law, and remembering that in the limit of high conductivity, such as the ISM, the only electric fields present are those that are induced by motions in the gas with velocity $v$ ($\ll c$), and hence the second term on the right in Ampère's law can be ignored, simplifying it to the magnetostatic case

$$\nabla \times \mathbf{B} = \frac{4\pi}{c} \mathbf{J}. \tag{3.29}$$

Substituting into the force equation we have

$$\mathbf{f}_{\mathrm{m}} = \frac{1}{4\pi} \left( \nabla \times \mathbf{B} \right) \times \mathbf{B} = -\frac{1}{4\pi} \mathbf{B} \times \left( \nabla \times \mathbf{B} \right). \tag{3.30}$$

We can simplify this equation using the vector identity

$$\frac{1}{2} \nabla \left( \mathbf{B}.\mathbf{B} \right) = \left( \mathbf{B}.\nabla \right) \mathbf{B} + \mathbf{B} \times \left( \nabla \times \mathbf{B} \right), \tag{3.31}$$

and thus

$$\mathbf{B} \times (\nabla \times \mathbf{B}) = \frac{1}{2} \nabla (\mathbf{B}.\mathbf{B}) - (\mathbf{B}.\nabla) \mathbf{B}. \tag{3.32}$$

Then the force equation becomes

$$\mathbf{f}_{\mathrm{m}} = -\nabla \left( \frac{B^2}{8\pi} \right) + \frac{1}{4\pi} (\mathbf{B}.\nabla) \mathbf{B}. \tag{3.33}$$

We can now write a new version of the equation of motion, also accounting for magnetic forces (in which, to preserve generality, we employ the gravitational potential $\Phi$)

$$\rho \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = -\nabla p - \rho \nabla \Phi - \nabla \left( \frac{B^2}{8\pi} \right) + \frac{1}{4\pi} (\mathbf{B}.\nabla) \mathbf{B}. \tag{3.34}$$

We can obtain an intuitive physical feel for the meaning of the magnetic terms in equation 3.34. Notice that the first magnetic term on the right-hand side enters in an identical way to the gas pressure. Hence this term is known as the *magnetic pressure*. Any gradient in the magnetic pressure results in a net force on the fluid, just as in the case of the gas pressure. Hence a region of the ISM with a high $\mathbf{B}$ value tends to be over-pressured relative to neighboring regions with lower $\mathbf{B}$, and will tend to expand.

The second magnetic term on the right-hand side of equation 3.34 can be understood as follows. We can write $\mathbf{B} = B\mathbf{s}$, where $\mathbf{s}$ is a unit vector in the direction of the field. Then we have

$$(\mathbf{B}.\nabla) \mathbf{B} = B\mathbf{s} \frac{\mathrm{d}(B\mathbf{s})}{\mathrm{d}x} = B^2 \mathbf{s} \frac{\mathrm{d}\mathbf{s}}{\mathrm{d}x} + Bs^2 \frac{\mathrm{d}B}{\mathrm{d}x}, \tag{3.35}$$

where $x$ is the direction along the field. For a constant magnetic field, $dB/dx$ is zero and the second term vanishes. Furthermore, if the field lines are straight, then $d\mathbf{s}/dx$ is zero and the first term vanishes. Hence this term clearly relates to how 'bent' the magnetic field lines are. It is found that the more bent the field lines, the stronger the restoring force. The term is therefore sometimes referred to as the *magnetic tension*.

## 3.2 Virial theorem

More force terms may be added to the equation of motion 3.34, for instance due to viscous processes[3], but we will leave it at this. The momentum equation describes the local behavior of a fluid, but it can be used to derive a global dynamical condition. This condition is referred to as the *virial theorem* and it is an extremely powerful mathematical tool. The word 'virial' is derived from the Latin 'vis' meaning 'energy' and was introduced in 1870 by the Polish-German physicist Rudolph Clausius ($1822-1888$).

We consider a system bounded by a closed surface $S$. The surface $S$ is assumed to move with the local fluid velocity $v$, so that there is no mass flow across the surface: the mass within $S$ is

---

[3]A fluid having no or negligible viscosity is named an inviscid fluid.

conserved. The pressure $p$ and the magnetic field $\mathbf{B}$ are not required to vanish at $S$, but we will assume that the material exterior to $S$ contributes negligibly to the gravitational acceleration $\nabla \Phi$. Essentially the proof of the virial theorem relies on the *moment of inertia*

$$I = \int_V \rho \, r^2 \, dV, \tag{3.36}$$

and requires integration over the entire volume $V$ enclosed by $S$. The virial theorem then states[4]

$$\frac{1}{2} \frac{d^2 I}{dt^2} = 2E_{\text{kin}} + 2E_{\text{th}} + E_{\text{mag}} + E_{\text{grav}}, \tag{3.37}$$

where

$$E_{\text{kin}} = \frac{1}{2} \int_V \rho \, v^2 \, dV, \tag{3.38}$$

$$E_{\text{th}} = \frac{3}{2} \int_V p \, dV, \tag{3.39}$$

$$E_{\text{mag}} = \int_V \frac{B^2}{8\pi} \, dV, \qquad \text{and} \tag{3.40}$$

$$E_{\text{pot}} = \int_V -\frac{1}{2} \sum_i \sum_{j \neq 1} \frac{G m_i m_j}{|\mathbf{r}_i - \mathbf{r}_j|} = -\frac{G}{2} \int_V dV_1 \int_V dV_2 \frac{\rho(\mathbf{r_1})\rho(\mathbf{r_2})}{|\mathbf{r_1} - \mathbf{r_2}|}. \tag{3.41}$$

$E_{\text{kin}}$ represents the total kinetic energy of the gas in bulk motion; $E_{\text{th}}$ is the energy contained in random, thermal motions, $E_{\text{mag}}$ is the magnetic energy, and $E_{\text{grav}}$ is the gravitational potential energy. The factor 1/2 in the equation for the potential energy enters because each mass $m$ or density element $\rho$ appears twice in the double sum. If the pressure and magnetic field have uniform values $p_\circ$ and $B_\circ$ at the bounding surface $S$, then the surface integrals can be evaluated, and the virial theorem becomes

$$\frac{1}{2} \frac{d^2 I}{dt^2} = 2E_{\text{kin}} + 2\left(E_{\text{th}} - E_{\text{th},\circ}\right) + \left(E_{\text{mag}} - E_{\text{mag},\circ}\right) + E_{\text{grav}}, \tag{3.42}$$

where

$$E_{\text{th},\circ} = \frac{3}{2} p_\circ V, \tag{3.43}$$

$$E_{\text{mag},\circ} = \frac{B_\circ^2}{8\pi} V. \tag{3.44}$$

The usual application of the virial theorem is to systems that are in a steady state, periodic or oscillatory, with time-average $\langle d^2 I / dt^2 \rangle = 0$. Then

$$2\langle E_{\text{kin}} \rangle + 2\langle E_{\text{th}} - E_{\text{th},\circ} \rangle + \langle E_{\text{mag}} - E_{\text{mag},\circ} \rangle + \langle E_{\text{grav}} \rangle = 0. \tag{3.45}$$

---

[4]For a derivation, see for instance *Physics of the interstellar and Intergalactic Medium*, by Bruce T. Draine, section 35.5 (pages 395-396) and appendix J (pages 508-510).

The condition of equilibrium for a stable, gravitationally bound system, with negligible contributions from external (over or under) pressure, magnetic fields, turbulence, and bulk motions such as rotation is therefore

$$2 \langle E_{\text{th}} \rangle + \langle E_{\text{grav}} \rangle = 0. \tag{3.46}$$

If twice the total internal kinetic energy of a molecular cloud exceeds the absolute value of the gravitational potential energy, the force due to the gas pressure will dominate the force of gravity and the cloud will expand. Conversely, if the internal kinetic energy is too low, the cloud will collapse. The boundary between these two cases describes the critical condition for stability.

Note that the energy in random, thermal motion $E_{\text{th}} = (3/2) \, N \, k \, T$, where $N$ is the total number of particles in a given volume $V$, is equivalent to

$$E_{\text{th}} = \frac{3}{2} N \, kT = N \, \frac{1}{2} \, m \, \langle v_{\text{th}}^2 \rangle \tag{3.47}$$

where $m$ is the mass of a gas particle and $\langle v_{\text{th}}^2 \rangle$ is the mean square thermal speed of the particles.

### Total energy

The condition (3.46) may also be expressed in terms of the total energy of the system by using the relations $\langle E \rangle = \langle E_{\text{th}} \rangle + \langle E_{\text{grav}} \rangle$. Thus

$$\langle E \rangle = \frac{1}{2} \langle E_{\text{grav}} \rangle . \tag{3.48}$$

The virial theorem applies to a wide variety of systems, from an ideal gas, to a binary orbit, to a gas cloud, to a cluster of stars or even a cluster of galaxies. For instance, consider the case of a static star. In equilibrium a star must obey the virial theorem, implying that its total energy is negative, one-half of the total potential energy. Assuming that the star formed as a result of the gravitational collapse of a large cloud, the potential energy of the system must have changed from an initial value of nearly zero to its negative static value. This implies that the star must have lost energy in the process, meaning that (half of the) gravitational energy must have been radiated into space during the collapse.

### Support of Giant Complexes

It is interesting to estimate how, in typical conditions prevailing in molecular clouds, the diffusive terms $E_{\text{kin}}$, $E_{\text{th}}$, and $E_{\text{mag}}$ compare with the attractive term $E_{\text{grav}}$. Let us assume a cloud with mass $M$, radius $R$, temperature $T$ and mean molecular weight $\mu$. Expressed in these quantities, the gravitational potential is $|E_{\text{grav}}| = (3/5) \, G \, M^2/R \sim G \, M^2/R$. For the thermal energy it follows that $E_{\text{th}} = (3/2) M \, \mathcal{R} \, T/\mu \sim M \, \mathcal{R} \, T/\mu$. The magnetic energy is approximated by $(B^2/8\pi)(4\pi R^3/3) \sim B^2 R^3/6$. The bulk velocity within giant clouds stems

mostly from the random motion of their clumps. Let us denote $\Delta v$ to the mean value of this speed, such that $E_{\text{kin}} = (1/2)\, M\, \Delta v^2$. We then find that

$$\frac{E_{\text{th}}}{E_{\text{grav}}} \approx \frac{M\, \mathcal{R}\, T}{\mu}\frac{R}{G\, M^2} \approx 0.003 \times \left(\frac{M}{10^5\, M_\odot}\right)^{-1}\left(\frac{R}{25\,\text{pc}}\right)\left(\frac{T}{15\,\text{K}}\right) \tag{3.49}$$

$$\frac{E_{\text{mag}}}{E_{\text{grav}}} \approx \frac{B^2\, R^3}{6}\frac{R}{G\, M^2} \approx 0.9 \times \left(\frac{B}{20\,\mu\text{G}}\right)^2\left(\frac{R}{25\,\text{pc}}\right)^4\left(\frac{M}{10^5\, M_\odot}\right)^{-2} \tag{3.50}$$

$$\frac{E_{\text{kin}}}{E_{\text{grav}}} \approx \frac{1}{2}M\, \Delta v^2\frac{R}{G\, M^2} \approx 0.5 \times \left(\frac{\Delta v}{4\,\text{km s}^{-1}}\right)^2\left(\frac{M}{10^5\, M_\odot}\right)^{-1}\left(\frac{R}{25\,\text{pc}}\right) \tag{3.51}$$

In the evaluations of these expressions, the typical size $R$ of a cloud follows in a straightforward way from the apparent size and the distance, the mass $M$ from the emission measure of optically thin diagnostic lines and some assumptions, the velocity dispersion $\Delta v$ from the Doppler width of spectral lines, generally one of CO, the temperature $T$ from the excitation of molecular lines, and the magnetic field $B$ from the Zeeman splitting of either the 21 cm line of H I or of a cluster of OH lines near 18 cm.

The important thing to notice is that the contribution of the gas to the total pressure is the lowest, and in fact negligible as a supporting factor of a molecular cloud. The two other terms are however sizable, and in magnitude of the order of the gravitational term. Both the internal 'turbulent' motions and the magnetic field pressure within a cloud are good candidates to provide the support against gravity within (giant) molecular clouds, and the discussion in the literature on which of both dominates is still very lively.

### No support of Giant Complexes

Let us briefly discuss what would happen if a molecular cloud is not supported at all against gravity. In that case equation (3.42) reduces to

$$\frac{1}{2}\frac{\text{d}^2 I}{\text{d}t^2} = E_{\text{grav}}, \tag{3.52}$$

i.e., only the potential energy will feature in the virial equation and we should not assume equilibrium, as for sure the cloud will collapse. Let us limit ourselves to a dimensional analysis of a homogeneous sphere of radius $R$ and mass $M$. The moment of inertia of such a sphere is $I = (2/5)\, M R^2$ and $|E_{\text{grav}}| = (3/5)\, G M^2/R$. Dimensional analysis then yields

$$\frac{1}{5}\frac{M R^2}{\tau^2} = \frac{3}{5}\frac{G M^2}{R}, \tag{3.53}$$

where $\tau$ is the characteristic timescale for free-fall collapse. Isolating $\tau$ and introducing the mean density yields

$$\tau = \left(\frac{1}{4\pi}\frac{1}{G\, \rho}\right)^{1/2}. \tag{3.54}$$

This is within a factor of two equal to the formal definition of the free-fall timescale $\tau_{\text{ff}}$ (equation 3.78), that will be discussed in section 3.4.

### 3.3 Gravitational instability in a static, uniform, non-magnetic medium

If one or (many) more stars are to form in a molecular cloud, then the cloud must become gravitationally unstable, and subsequently collapse. In this section we consider the question of a cloud's stability. The answer, we will see below, depends upon the size of the spherical portion. For a large portion the self-gravity overcomes internal pressure, while for a small portion the internal pressure resists gravity.

The problem of the gravitational stability of a self-gravitating cloud has first been addressed by the British astronomer James Jeans ($1877-1946$). In this traditional picture, the only restoring force is the gas pressure gradient, linked to the thermal energy of the cloud. The considerations presented in section 3.2 show that such a picture is a poor approximation for the first phases of the contraction, but it must have some validity, since the gas pressure gradient (for massive stars aided by radiation pressure gradients) ends up as the dominant restoring force in the end products, i.e. stars. The criterium for collapse obtained by Jeans can be found by linearly perturbing the relevant equations; we here pursue an approach using the virial theorem.

Consider an infinite, non-magnetic, static medium, with initial uniform density $\rho$ and uniform isothermal temperature $T$, hence isothermal sound speed $a = \sqrt{kT/\mu m_\mathrm{H}}$ (see equation 3.3). We focus on a spherical volume in this medium, with radius $R$ and mass $M = \mu\, m_\mathrm{H}\, N$, where $\mu$ is the mean molecular weight in units of $m_\mathrm{H}$ and $N$ is the total number of particles in the volume. The gravitational potential energy is

$$E_\mathrm{grav} = -\frac{3}{5}\frac{G\, M^2}{R}. \tag{3.55}$$

The cloud's internal kinetic energy, i.e. that contained in random, thermal motions, using equation 3.47, is given by

$$E_\mathrm{th} = \frac{3}{2}\, N\, k\, T = \frac{3}{2}\frac{M\, kT}{\mu\, m_\mathrm{H}} \tag{3.56}$$

Now, by the virial theorem equation 3.46, the condition for collapse becomes

$$\frac{3\, M\, kT}{\mu\, m_\mathrm{H}} < \frac{3}{5}\frac{G\, M^2}{R}. \tag{3.57}$$

The radius may be replaced by using the initial mass density of the cloud, $\rho$, assumed here to be constant throughout the cloud,

$$R = \left(\frac{3M}{4\pi\, \rho}\right)^{1/3}. \tag{3.58}$$

This yields the so-called *Jeans instability criterion*,

$$M > M_\mathrm{J} \;=\; \left(\frac{5k\, T}{G\, \mu\, m_\mathrm{H}}\right)^{3/2}\left(\frac{3}{4\pi\, \rho}\right)^{1/2} = \left(\frac{5k\, T}{G\, \mu\, m_\mathrm{H}}\right)^{3/2}\left(\frac{3}{4\pi\, \mu\, m_\mathrm{H}\, n}\right)^{1/2} \tag{3.59}$$

$$= 93.7 \frac{T^{3/2}}{\mu^2 \, n^{1/2}} \, M_\odot \tag{3.60}$$

$$= 7.41 \left(\frac{T}{10\,\mathrm{K}}\right)^{3/2} \left(\frac{2}{\mu}\right)^2 \left(\frac{10^4}{n}\right)^{1/2} M_\odot \tag{3.61}$$

where $M_\mathrm{J}$ is the *Jeans mass*, i.e. the minimum mass necessary to initiate the spontaneous collapse of the cloud. The particle density $n$ is per $\mathrm{cm}^3$.

Using equation 3.58, the Jeans criterion may also be expressed in terms of the minimum radius $R_\mathrm{J}$ necessary to collapse a cloud of density $\rho$. For collapse to occur it is required that

$$R > R_\mathrm{J} = \left(\frac{15 k \, T}{4\pi \, G \, \mu \, m_\mathrm{H} \, \rho}\right)^{1/2} = \left(\frac{15 k \, T}{4\pi \, G \, \mu^2 \, m_\mathrm{H}^2 \, n}\right)^{1/2} \tag{3.62}$$

$$= 9.7 \frac{1}{\mu} \left(\frac{T}{n}\right)^{1/2} \mathrm{pc,} \tag{3.63}$$

where $R_\mathrm{J}$ is the *Jeans length*. The particle density $n$ is per $\mathrm{cm}^3$.

To get some feel for the orders of magnitude of the values of the Jeans length and mass encountered in studies of molecular clouds, we can insert some typical values into equations 3.61 and 3.63. The densest regions of molecular clouds in which no stars have yet formed are typically observed to have temperatures of around 20 K, particle densities of around $10^5$ hydrogen molecules per $\mathrm{cm}^3$, and a mean molecular weight of 2.3 (i.e. mainly molecular hydrogen with atomic helium). For such a region we derive a Jeans length of about 0.06 parsec and a Jeans mass of about 5 $M_\odot$. So if, for instance, we observed a uniform density region at 20 K which was 0.1 pc in diameter, and contained 5 $M_\odot$ of matter, we would note that this was greater than its Jeans mass. Therefore, we would say that it was Jeans-unstable, and we would predict that it was about to collapse to form a star.

## The hierarchical fragmentation of collapsing clouds

We can also consider a giant molecular cloud as a whole. With a typical density of $n = 100\,\mathrm{cm}^{-3}$ and $T = 15\,\mathrm{K}$ (see table 2.2) regions larger than $R_\mathrm{J} \simeq 2\,\mathrm{pc}$ would be prone to collapse (adopting $\mu = 2$). Typical sizes of GMCs are however much larger and such systems should thus be Jeans unstable. Given that the mass of a GMC is in the order of $10^5 \, M_\odot$, would this collapse lead to the formation of a $10^5 \, M_\odot$ star? Observations show that this does not happen. One reason is that GMCs are actually inhomogeneous and clumped on all scales, as witnessed by the clouds shown in the images in chapter 2; if anything, the complexes themselves appear to be swarms of more coherent clumps, suggesting that the collapse of cold neutral gas from which the GMCs originate starts a *hierarchical fragmentation*[5] that is

---

[5]In hierarchical fragmentation sub-regions in the cloud break-up that later may again fragment in even smaller sub-units, etcetera.
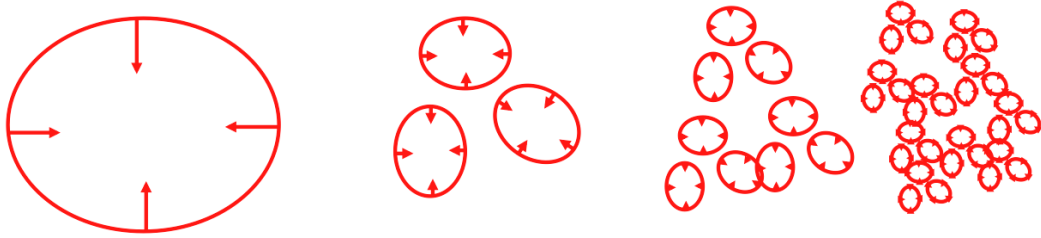
Figure 3.1: *Schematic representation of the hierarchical fragmentation of collapsing molecular clouds. As the density increases during collapse, the Jeans mass decreases, causing progressive fragmentation. The fragmentation stops when the clumps are so dense that they become opaque for their own radiation, causing the cloud fragment to heat up. This increases the Jeans mass, effectively halting further fragmentation.*

still ongoing in the giant molecular clouds. For the record, the collapse of a homogeneous medium of cold atomic gas, with $T = 100$ K, $\mu = 1.3$, and $n = 1 - 10 \, \mathrm{cm}^3$ corresponds to $M_\mathrm{J} \simeq (2 - 6) \times 10^5 \, M_\odot$ and $R_\mathrm{J} \simeq (25 - 75)$ pc, i.e. quantities typical for GMCs.

Digressing a bit from our line of reasoning, the clumps in giant-molecular clouds seem to be self-gravitating. Employing typical clump parameters of $n = 10^3 \, \mathrm{cm}^{-3}$ and $T = 10$ K in equation 3.61, we find that $M_\mathrm{J} \simeq 25 \, M_\odot$, an order of magnitude below the actual observed masses. To put the matter another way, the internal temperature of a $\sim 250 \, M_\odot$ clump would have to be about 50 K for it to be supported entirely by thermal pressure. Since measured temperatures are much lower, at least throughout most of the interior, and since the clumps are apparently *not* undergoing global collapse, one or more extra sources of support are necessary. The most plausible sources are the interstellar magnetic field and large scale turbulence.

We return to our discussion of collapse based on the Jeans equations. The process of fragmentation that segments a collapsing cloud is an aspect of star formation that receives significant attention. An important consequence of the collapse of a molecular cloud is that the density increases by many orders of magnitude (during free-fall). Consequently, since $T$ remains nearly constant throughout much of the collapse, it appears that the Jeans mass must decrease as $M_\mathrm{J} \propto n^{-1/2}$. After collapse has begun, any initial inhomogeneities in density will cause individual sections of the cloud to satisfy the Jeans mass limit independently and begin to collapse locally, producing smaller features within the original cloud. This cascading collapse could lead to the formation of large numbers of smaller objects.

What is it that stops the fragmentation process? The answer to the question lies in our implicit assumption that the collapse is isothermal. If the energy that is released during a graviational collapse is radiated away efficiently, the temperature can remain nearly constant. If the cloud starts to become opaque for radiation, energy can not easily escape any longer and the temperature must rise. In the most extreme case energy can not be transported out of the cloud at all, i.e. the collapse is adiabatic. Of course, the real situation must be somewhere between these two limits, but by considering each of these special cases, we can start to understand some of

the important aspects of the problem.

Using the ideal gas law, equation 3.2, an adiabatic relation between density and temperature can be obtained from equation 3.4,

$$T = K' \mu \rho^{\gamma-1}, \tag{3.64}$$

where $K'$ is a constant. Substituting this expression into equation 3.61, we find that for an adiabatic collapse, the dependence of the Jeans mass on density becomes

$$M_{\mathrm{J}} \propto \rho^{(3\gamma-4)/2}. \tag{3.65}$$

For atomic hydrogen $\gamma = 5/3$, giving $M_{\mathrm{J}} \propto \rho^{1/2} \propto n^{1/2}$; the Jeans mass *increases* with increasing density for a perfectly adiabatic collapse of a cloud. This behavior means that the collapse results in a minimum value for the mass of the fragments produced. The minimum mass depends on the point when the collapse goes from being predominantly isothermal to adiabatic.

Of course, this transition is not instanteneous or even complete. However, it is possible to make a crude order-of-magnitude estimate of the lower mass limit of the fragments. As we have already mentioned in the discussion of the virial theorem (see 'Total energy' paragraph of section 3.2) energy must be liberated during the collapse of the cloud. The energy released is roughly

$$\Delta E_{\mathrm{grav}} \simeq \frac{3}{10} \frac{G M_{\mathrm{J}}^2}{R_{\mathrm{J}}} \tag{3.66}$$

for a spherical cloud just satisfying the Jeans criterion at some point during the collapse. Averaging over the characteristic time it takes for a cloud to collapse, the free-fall time (see section 3.4), yields for the average luminosity due to liberation of gravitational energy

$$L_{\mathrm{ff}} \simeq \frac{\Delta E_{\mathrm{grav}}}{\tau_{\mathrm{ff}}} = \frac{3.2}{\pi} \left(\frac{3}{4\pi}\right)^{1/2} G^{3/2} \left(\frac{M_{\mathrm{J}}}{R_{\mathrm{J}}}\right)^{5/2} \tag{3.67}$$

If the cloud were optically thick and in thermodynamic equilibrium, the energy would be emitted as blackbody radiation. However, because of extinction of radiation by in-falling material in higher-up optically thinnish layers the radiation losses are less efficient than for an ideal blackbody. We express the radiated luminosity as

$$L_{\mathrm{rad}} = f \, 4\pi R^2 \, \sigma T^4 \tag{3.68}$$

where an efficiency factor, $0 < f < 1$, has been introduced to indicate the deviation from thermodynamic equilibrium. If, relatively early-on in the collapse, escaping radiation does not interact at all with overlying in-falling material, $f \sim 0$. If, later-on in the collapse, energy emitted by some parts of the cloud is absorbed and then re-emitted by other parts of the cloud, thermodynamic equilibrium would more nearly apply and $f$ would be closer to unity. Equating the two expressing for the cloud's luminosity, i.e. $L_{\mathrm{ff}} = L_{\mathrm{rad}}(R_{\mathrm{J}})$, we arrive at

$$M_{\mathrm{J}}^5 = \frac{64\pi^5}{3(3.2)^2} \frac{\sigma^2}{G^3} f^2 \, T^8 \, R_{\mathrm{J}}^9. \tag{3.69}$$

Using equation 3.58 to eliminate the radius, $\rho = \mu\, m_{\mathrm{H}}\, n$ to eliminate the density, and equation 3.61 to express the particle number density $n$ in the Jeans mass, we arrive at an estimate of when adiabatic effects become important, expressed in terms of the minimum obtainable Jeans mass:

$$M_{J_{\min}} = 0.02\, \frac{T^{1/4}}{\mu^{9/4}\, f^{1/2}}\, M_{\odot}. \tag{3.70}$$

If we take $\mu \sim 1$, $f \sim 0.1 - 1$, and $T \sim 1000\,\mathrm{K}$ at the time when adiabatic effects may start to become significant, we find $M_{J_{\min}} \sim 0.1 - 0.3\, M_{\odot}$, i.e. fragmentation ceases when the segments of the original cloud begin to reach the range of sub-solar mass objects. Indeed, this is very different from the $\sim 10^5\, M_{\odot}$ star that would have formed if hierarchical fragmentation would not have occured.

## 3.4 Pressure-free collapse

In the case that the criterion for gravitational collapse has been satisfied in the absence of rotation, turbulence, and/or magnetic fields, the molecular cloud will collapse. If we make the simplifying assumption that any existing pressure gradients are too small to influence the motion appreciably, then the cloud is essentially in *free-fall* during the first part of its evolution. We furthermore assume that throughout the free-fall phase the temperature of the gas remains constant, i.e. the collapse is said to be isothermal. The latter assumption can be justified as long as the cloud remains optically thin and the gravitational potential energy released during the collapse can be efficiently radiated away. The timescale associated with this free-fall collapse phase is the *free-fall timescale* $\tau_{\mathrm{ff}}$. We shall derive an expression for $\tau_{\mathrm{ff}}$ below. You should notice that the free-fall time is actually independent of the initial radius of the sphere. Consequently, as long as the original density of the spherical molecular cloud was uniform, all parts of the cloud will take the same amount of time to collapse, and the density will increase at the same rate everywhere. This behavior is known as a *homologous collapse*.

However, if the cloud is somewhat centrally condensed when the collapse begins, the free-fall time will be shorter for material near the center than for material farther out. Thus as the collapse progresses, the density will increase more rapidly near the center than in other regions. In this case the collapse is referred to as an *inside-out collapse*.

To evaluate $\tau_{\mathrm{ff}}$ it is convenient to consider the Lagrangian description in which coordinates move with the flow. Euler equation 3.23, using Eq. 3.19, in radial coordinates for a pressure-less situation reduces to

$$\frac{\mathrm{d}v}{\mathrm{d}t} = \frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = -\frac{GM(r)}{r^2}. \tag{3.71}$$

Of course the right-hand side is just the local acceleration of gravity at a distance $r$ from the center of a spherical cloud. As usual, the mass of the sphere interior to the radius $r$ is denoted by $M(r)$.

To describe the behavior of the surface of a sphere of radius $r$ within the collapsing cloud as a function of time, equation 3.71 must be integrated over time. Since we are interested only in the surface that encloses $M(r)$, the mass interior of $r$ will remain a constant during collapse, i.e. $M(r) = M$. If we multiply both sides by the velocity of the surface of the sphere, we arrive at the expression

$$\frac{\mathrm{d}r}{\mathrm{d}t} \frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = -\frac{GM}{r^2} \frac{\mathrm{d}r}{\mathrm{d}t}, \tag{3.72}$$

which can be integrated with respect to time to give

$$\frac{1}{2} \left( \frac{\mathrm{d}r}{\mathrm{d}t} \right)^2 = \frac{GM}{r} + K. \tag{3.73}$$

The integration constant $K$ can be evaluated by requiring that the velocity of the sphere's surface be zero at the beginning of the collapse, or $dr/dt = 0$ when $r = r_\circ$. This gives $K = -GM/r_\circ$. Realizing also that $M = 4\pi r_\circ^3 \rho_\circ / 3$ we obtain

$$\frac{\mathrm{d}r}{\mathrm{d}t} = - \left[ \frac{8\pi}{3} G \rho_\circ r_\circ^2 \left( \frac{r_\circ}{r} - 1 \right) \right]^{1/2}. \tag{3.74}$$

Note that the negative root was chosen because the cloud is collapsing. To facilitate the integration we introduce the normalized coordinate $x = r/r_\circ$ and the constant $\chi = (8\pi G \rho_\circ / 3)^{1/2}$, which leads to the differential equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\chi \left( \frac{1}{x} - 1 \right)^{1/2}. \tag{3.75}$$

Making yet another substitution $x = \cos^2 \theta$, and after some manipulation, we arrive at

$$\cos^2 \theta \frac{\mathrm{d}\theta}{\mathrm{d}t} = \frac{\chi}{2}. \tag{3.76}$$

This equation may be integrated directly with respect to $t$ to yield

$$\frac{\theta}{2} + \frac{1}{4} \sin 2\theta = \frac{\chi}{2} t + C. \tag{3.77}$$

Lastly, the integration constant $C$ must be evaluated. Using that $r = r_\circ$ at $t = 0$, implying that $x = 1$ or $\theta = 0$ at the beginning of the collapse, it follows that $C = 0$. Let $t = \tau_\mathrm{ff}$ when the radius of the collapsing sphere reaches zero, i.e. $x = 0$, or equivalently $\theta = \pi/2$, then $t_\mathrm{ff} = \pi/(2\chi)$. Substituting the value for $\chi$, we have

$$\tau_\mathrm{ff} = \left( \frac{3\pi}{32} \frac{1}{G \rho_\circ} \right)^{1/2} \tag{3.78}$$

$$= 3.4 \times 10^5 \left( \frac{10^4 \, \mathrm{cm}^{-3}}{n} \right)^{1/2} \quad \mathrm{yr}, \tag{3.79}$$

where $\rho = \mu \, m_\mathrm{H} \, n$ is used for the last equality, and $\mu = 2.3$ is adopted. As announced at the start of the derivation, because this result depends only on $\rho_\circ$, and not on the value of the starting radius $r_\circ$, *all* fluid elements reach $r = 0$ at precisely the same instant.

## Exercise 3.1

Thermal energy, also known as random or internal kinetic energy, is the result of vibrational, rotational, and translational motion of the particles in the system. The equipartition theorem states that in thermal equilibrium the average energy of each of the $f$ degrees of freedom (each independent way the system can move) is $kT/2$, hence, for a single particle $E_{\text{th}} = (f/2)\, kT$.

a) Show that for a volume $V$ of constant particle density $n$ and temperature $T$, thermal pressure and thermal energy are related as

$$p = (\gamma - 1)\frac{E_{\text{th}}}{V}, \tag{3.80}$$

where $\gamma$ is the adiabatic index (see Eq. 3.5).

b) Explain heuristically why one indeed should expect pressure to be proportional to energy divided by volume.

Notice that in Eq. 3.39 we have tacitly assumed the gas to be atomic, i.e. $f = 3$.

## Exercise 3.2

The moment of inertia must be specified with respect to a chosen axis of rotation. For a point mass the moment of inertia is just the mass times the square of the perpendicular distance to the rotation axis, $I = m\, r^2$. That point mass relationship becomes the basis for all other moments of inertia since any object can be built up from a collection of $N$ point masses

$$I = \sum_{i=1}^{N} m_i\, r_i^2 \tag{3.81}$$

For a continuous distribution of mass the summation is replaced by an integral and $I$ may be written as

$$I = \int_M r^2\, dm = \int_V \rho(r)\, r^2\, dV, \tag{3.82}$$

where $M$ is the mass and $V$ the volume of the body.

a) Compute the moment of inertia of a homogeneous flat disk of density $\rho$, height $h$ and radius $R$, and express $I$ as a function of $M$ and $R$. The axis of rotation is through the center of the disk and perpendicular to the disk surface.

b) Compute the moment of inertia of a homogeneous sphere of density $\rho$ and radius $R$ and express $I$ as a function of $M$ and $R$. The axis of rotation is through the center of the sphere.

**Exercise 3.3**

a) The angular momentum of a mass rotating at a distance $r$ from the rotation axis (i.e. the $r$ from cylindrical coordinates) at angular speed $\omega$ is given by

$$J = m\,\omega\,r^2 = m\,v\,r \tag{3.83}$$

Calculate the angular momentum contained in the orbits of Earth, Jupiter, Saturn, and Neptune. Also compute the specific angular momentum $j$, i.e. the angular momentum per unit mass.

b) Calculate the angular momentum and specific angular momentum of the Sun. You can assume that the Sun is a spherical rigid body rotator.

**Exercise 3.4**

Calculate the angular momentum per unit mass of a molecular cloud with a mass of one solar mass and a radius of 0.1pc. Assume rigid rotation, with $\omega = 10^{-14}$ rad s$^{-1}$. Consider two cases:

a) constant mass density (because it is simple), and

b) a mass density proportional to $r^{-2}$ (because that is a realistic initial condition before cloud collapse).

**Exercise 3.5**

We treat a cloud as an ensemble of particles, having mass $m_i$, position $\mathbf{r}_i$, and velocity $\mathbf{v}_i$, moving in their mutual gravitational field. At the bounding surface of the cloud the external pressure is zero. Magnetic fields are absent. The moment of inertia of the cloud is given by

$$I = \sum_i m_i\,\mathbf{r}_i \cdot \mathbf{r}_i. \tag{3.84}$$

In the case of a cloud in equilibrium, the time derivative and the second time derivative of the moment of inertia must be zero. Prove virial theorem Eq. 3.37.

**Exercise 3.6**

Consider a spherical isothermal cloud at temperature $T$, mean molecular weight $\mu$, mass $M$ and radius $R$, and external surface pressure $p_\circ$. The isothermal sound speed $a$. We assume that the sphere has uniform density. The virial equilibrium condition relevant for this situation is

$$2E_{\text{th}} - 3p_\circ V + E_{\text{grav}} = 0, \tag{3.85}$$

where $E_{\text{th}}$ is the thermal energy and $E_{\text{grav}}$ is the potential energy.

a) Derive an expression for the surface pressure $p_\circ$ as a function of $R$, $M$ and $a$.

b) Show that if we keep $M$ and $a$ fixed and vary the radius $R$, the pressure $p_\circ$ has a maximum value

$$p_\circ = \text{constant}\,\frac{a^8}{G^3\,M^2}, \tag{3.86}$$

and derive and compute the constant.

c) Explain in words why, when varying $R$ but maintaining virial equilibrium, one finds a maximum in the external pressure.

## Exercise 3.7

We assume a spherical cloud of initial radius $R_\circ$ that is in virial equilibrium and supported by magnetic pressure only. The initial magnetic field strength is $B_\circ$. Once a cloud collapses, the medium is not compressed parallel to the magnetic field lines. Therefore the total magnetic flux threading a given surface area perpendicular to the direction of the magnetic field lines is a conserved quantity. Use this property of the magnetic field and derive an expression for the minimum mass $M_{\text{mcm}}$ that is needed to initiate the spontaneous collapse of the cloud by self-gravity, and express the result as a function of $B_\circ$ and $R_\circ$. This minimum mass is referred to as the *magnetic critical mass*.

## Exercise 3.8

Derive the free-fall time, equation (3.78), for pressure-free collapse.

## Exercise 3.9

a) Start with eq. (3.11) and show that in planar geometry the time-dependent continuity equation is given by

$$\frac{1}{\rho}\frac{\partial\rho}{\partial t} + \frac{\partial v_z}{\partial z} + \frac{v_z}{\rho}\frac{\partial\rho}{\partial z} = 0 \tag{3.87}$$

b) Start with eq. (3.11) and show that in spherical geometry the time-dependent continuity equation for radial flow is given by

$$\frac{1}{\rho}\frac{\partial \rho}{\partial t} + v_r \left[\frac{1}{v}\frac{\partial v}{\partial r} + \frac{2}{r} + \frac{1}{\rho}\frac{\partial \rho}{\partial r}\right] = 0, \qquad (3.88)$$

where $v$ is the velocity in the radial direction.

c) Start with eq. (3.15) and show that for a stationary radial flow the continuity equation is given by

$$\frac{1}{v}\frac{\partial v}{\partial r} + \frac{2}{r} + \frac{1}{\rho}\frac{\partial \rho}{\partial r} = 0 \qquad (3.89)$$

## Exercise 3.10

a) Assume that in a planar geometry the only force field present is $f_z$ and that this field is acting in the $z$-direction only. Start with eq. (3.23) and show that in this geometry the time-dependent momentum equation is given by

$$\rho\left[\frac{\partial v_z}{\partial t} + v_z\frac{\partial v_z}{\partial z}\right] = -\frac{\partial p}{\partial z} + f_z \qquad (3.90)$$

b) Assume that in a spherical geometry the only force field present is $f_r$ and that this field is acting along the radial direction only. Start with eq. (3.23) and show that in this geometry the time-dependent momentum equation for radial flow is given by

$$\rho\left[\frac{\partial v}{\partial t} + v_r\frac{\partial v}{\partial r}\right] = -\frac{\partial p}{\partial r} + f_r \qquad (3.91)$$

## Exercise 3.11

For general equations of state, the speed of sound is given by

$$v_s^2 = \frac{\partial p}{\partial \rho} \qquad (3.92)$$

where the derivative is taken with respect to adiabatic change. It follows from eq. (3.2) that for an isothermal medium and a constant mean molecular weight $\mu$, $v_s = \sqrt{\mathcal{R}T/\mu}$ (in these notes we then adopt the notation $v_s = a$). Assume that the only force field present is gravity, such that $f_r = \rho g$ with $g = -GM/r^2$. Use the results eq. (3.89) and eq. (3.91) and show that for an isothermal medium in which the mean molecular weight is constant, the equation of motion for a stationary radial flow can be written as

$$\frac{1}{v}\frac{\partial v}{\partial r} = \left[\frac{2a^2}{r} - \frac{GM}{r^2}\right] / \left[v^2 - a^2\right]. \qquad (3.93)$$

This equation has a singularity at the point where $v(r) = a$, where the denominator is zero and $\partial v/\partial r$ becomes infinite (which cannot be physical), unless the numerator is also zero there. (As this equation traces the change of velocity in one direction only, i.e. $r$, it is quite custom to replace the partial derivative $\partial/\partial r$ by the total derivative $d/dr$.)

# 4

# SPH simulation of the collapse of molecular clouds

In a sense, the discussion of pressure-free collapse (section 3.4) might be seen as a very simplistic theory of star formation. Let us devote a chapter by jumping to the other extreme of describing star formation and qualitatively and quantitatively discuss aspects of 3D numerical simulations of the collapse of $\sim 10^3 \, M_\odot$ molecular clouds leading to the formation of a stellar cluster containing hundreds of stars. The first calculation we discuss is by Bonnell, Bate & Vine (2003, MNRAS 343, 413), that is reported on in a paper that is fundamental – it is an important step toward developing a predictive theory of star formation – and exceptionally well written and accessible. Further aspects of this calculation are reported on in Bonnell & Bate (2005, MNRAS 362, 915). Note that this is not the only such calculation; see for instance Bate, Bonnell & Bromm (2003, MNRAS 339, 577), Bate (2009, MNRAS 392, 590; 2012 MNRAS 419, 3115; 2019 MNRAS 484, 2341), and Krumholz et al. (2012, ApJ 754, 71). With increasing computing power and improved numerical methods, ever smaller and smaller spatial scales can be resolved. We discuss Bate (2018, MNRAS 475, 5618), building on Bate (2012 MNRAS 419, 3115), presenting, for the first time, hydrodynamical simulations that resolve individual proto-stellar disks around the forming stars.

## 4.1 Aspects of the SPH simulation

The numerical technique applied by Bonnell et al. is that of *Smooth Particle Hydrodynamics* (SPH) with Sink Particles. The details of this technique are beyond the scope of these lecture notes, but see figure 4.7 for a short description as given by Zinnecker & Yorke (2007). Though numerical techniques and computer hardware have seen tremendous development in the past decades, still many assumptions need to be made in order to evolve a 1000 $M_\odot$ cloud over several free-fall time scales. In the Bonnell et al. simulation the most important assumptions are:

- ○ The simulation accounts for $5 \times 10^5$ SPH particles and starts with a spherical cloud of 1000 $M_\odot$ and a diameter of 1 pc. The medium outside of the cloud is void.

- ○ The gas is forced to remain isothermal at 10 K throughout the simulation, emulating the effect of efficient radiative cooling at the low gas densities present.

- ○ The cloud is initially supported by a random turbulent velocity field with total turbulent energy equal to the gravitational energy; the corresponding mean line-of-sight turbulent velocity is about $1.3 \, \mathrm{km \, s^{-1}}$, which is approximately Mach 6. The thermal energy is initially only 1 percent of the kinetic energy.

- ○ Feedback, both radiative and kinematic, from the newly formed stars is neglected.

- ○ Magnetic fields are non-existent.

- ○ Once fragmentation has produced a 'proto'-star, the constituent SPH particles are replaced with a single sink-particle. These sink particles, used in the simulation to follow the newly formed stars, interact only through gravitational forces and by accretion of gas particles that fall into their sink-radii. Sink-particle creation occurs when the densest gas particle (at a given time) and its roughly 50 neighbors are self-gravitating, sub-virial and occupy a region smaller then the sink-radius of 200 AU – which corresponds to the typical size of a proto-planetary disk. This requires a gas density $\rho \gtrsim 1.5 \times 10^{-15} \, \mathrm{gr \, cm^{-3}}$ and implies a minimum mass of the SPH particle of $\sim 0.1 \, M_\odot$.

- ○ Gas particles are accreted if they fall within a sink radius of a sink particle and are bound to it. In the case of overlapping sink radii, the gas particle is accreted by the sink particle to which it is most bound.

- ○ The gravitational forces between stars are smoothed on a scale of 160 AU or less, so that the widest binaries and multiple systems are unresolved, nor are circumstellar disks. This gravity softening also implies that stellar collisions are not included in the simulations.

Four snapshots of the star-formation process are shown in figure 4.1. In discussing aspects of the simulation we closely follow and use excerpts from the Bonnell, Bate, & Vine (2003) study. The initial evolution of the molecular cloud is due to the turbulent motions present in the gas. The supersonic turbulence leads to the development of shocks, producing filamentary structures. The shocks also remove kinetic energy (assumed to be radiated away), and thus remove turbulent support locally. The chaotic nature of the turbulence leads to higher-density regions in the filamentary structures which, if they become self-gravitating, collapse to form stars. Star formation occurs simultaneously at several different locations in the cloud. The stars that form first are in the highest-density gas, where the dynamical time-scale (i.e. the local free-fall time) is the shortest. Surrounding clumps with slightly lower gas density form additional stars. Both the stars and the residual gas are attracted by their mutual gravitational forces and fall towards each other. Gas dynamics dampen the infall velocities, allowing the systems to rapidly merge to form a high-density sub-cluster, containing five to several tens of
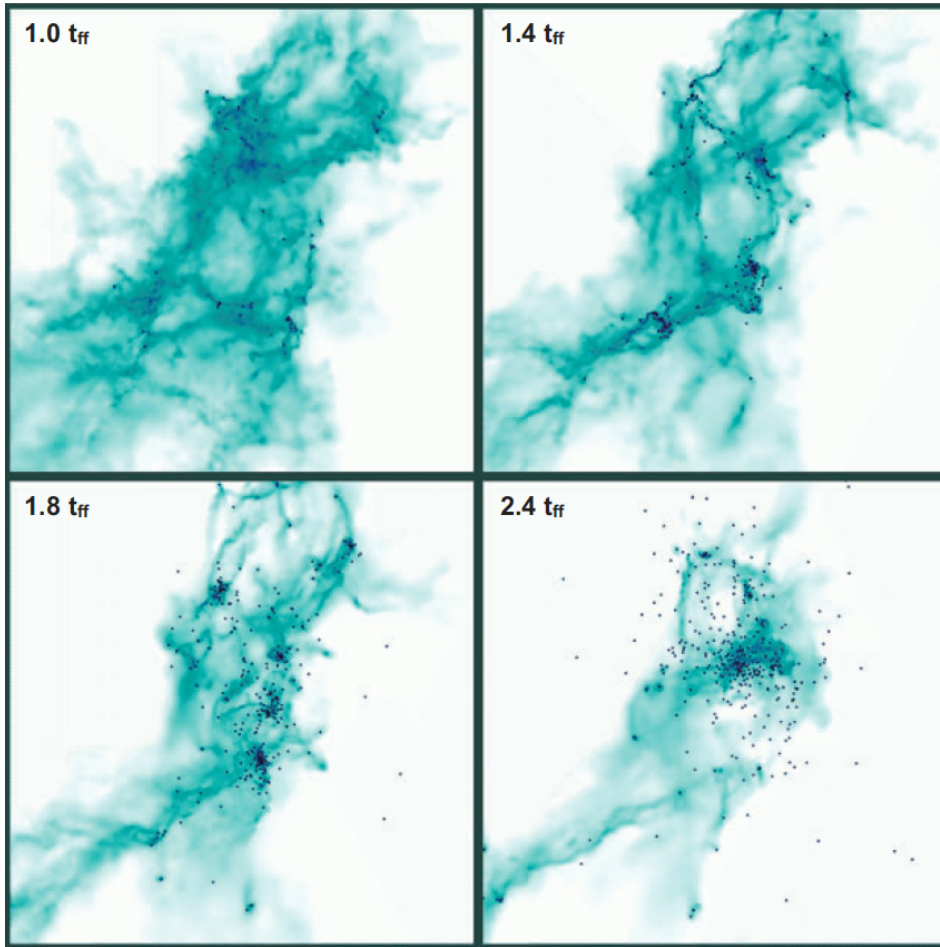
Figure 4.1: *SPH simulation of the collapse of a 1000 $M_\odot$ turbulent molecular cloud leading to the formation of a stellar cluster through hierarchical fragmentation. Each panel shows a region of 1 pc per side. The logarithm of the column density is plotted from a minimum of 0.025 (white) to a maximum of 250 (dark blue) $\mathrm{gr\,cm}^{-3}$. The stars are indicated by dark blue dots. The four panels capture the evolution of the system at times of 1.0, 1.4, 1.8 and 2.4 initial free-fall times, where the initial free-fall time for the cloud is $\tau_{\mathrm{ff}} = 1.9 \times 10^5$ yr. The turbulence causes shocks to form in the molecular cloud, dissipating kinetic energy and producing filamentary structures, which fragment to form dense cores and individual stars (at $1.0\,\tau_{\mathrm{ff}}$). The stars fall towards local potential minima and hence form subclusters (at $1.4\,\tau_{\mathrm{ff}}$). These subclusters evolve by accreting more stars and gas, ejecting stars, and by mergers with other subclusters (at $1.8\,\tau_{\mathrm{ff}}$). The final state of the simulation is a single, centrally condensed cluster with little substructure (at $2.4\,\tau_{\mathrm{ff}}$). The cluster contains more than 400 stars and has a gas fraction of approximately 16 per cent. From: Bonnell, Bate, & Vine, 2003, MNRAS 343, 413.*

stars. The number of stars in each sub-cluster increases as further star formation occurs nearby, and these stars fall into the existing potential wells. This process repeats itself until several hundred stars are formed and are mostly contained in a modest number of sub-clusters (in this simulation five such sub-clusters can be identified). The star formation activity decreases
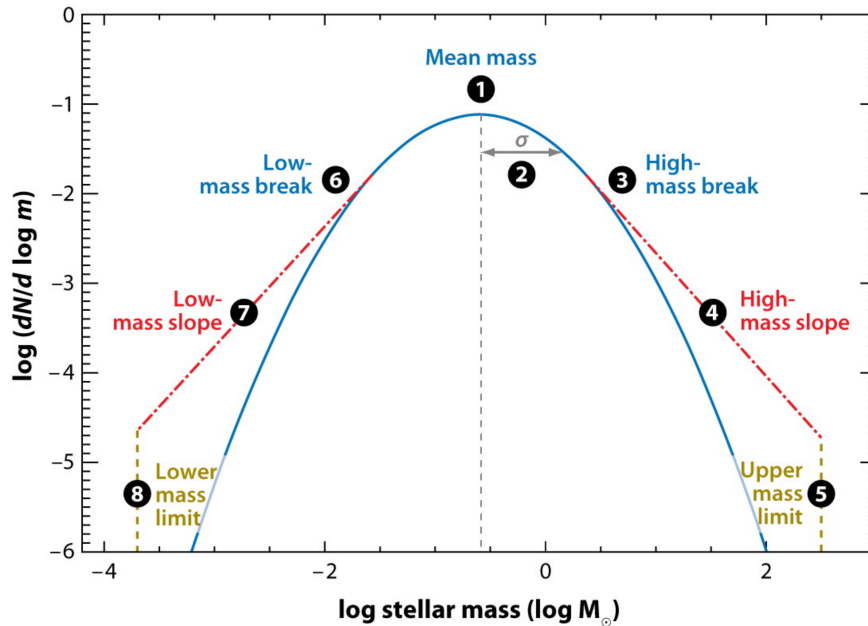
Figure 4.2: *The original work of Salpeter (1955) on determining the slope of the mass function considered stars in the solar neighborhood with masses in the range $0.4 - 10\,M_\odot$. He modeled this distribution using a power-law IMF (see Eq. 4.1). Of course, stars exist of both lower and higher mass. In the late 1970s it was recognized that over the full range of masses the IMF was probably not a single power law. A comprehensive description, using a log-normal distribution was first introduced by Miller & Scalo (1978), and a theoretical explanation for this functional form was offered by Zinnecker (1984; see also Larson 1973). The figure offers an overview of the (eight) free parameters of the log-normal IMF, in which the Salpeter law holds above some limit (high-mass break). Recent work by Schneider et al. (2018, Science 359, Issue 6371, 69) suggests that perhaps a 9th parameter is needed, as it appears that the IMF shows a 'kink' at masses above $10\text{-}15\,M_\odot$, flattening from a slope $\Gamma = 1.35$ to a slope $\Gamma = 0.9$. From: Bastian et al. 2010, ARA&A 48, 339.*

and the sub-clusters, aided by the dissipative effects of their embedded gas, sink towards each other and finally merge to form one single cluster containing over 400 stars. The final cluster is approximately spherical in shape with a centrally condensed core, as is observed in young stellar clusters. If we define the global stellar number-density by estimating the smallest volume required to contain half of the total number of stars, we find that at the end of the simulation, after $4.8 \times 10^5$ years, this number is $\sim 600$ stars pc$^{-3}$.

## 4.2 Initial mass function and star formation efficiency

In addition to acting as a reservoir for star formation and as a damping force of the stellar dynamics, gas is accreted on to individual stars, thereby increasing their masses. The stars

compete for the gas, with those in the bottom of their local potential wells accreting the most and becoming the most massive stars. In this simulation, gas accretion results in final stellar masses that range from ~0.07 to 27 $M_\odot$. The median and mean stellar mass are 0.43 $M_\odot$ and 1.38 $M_\odot$, respectively. The final mass distribution $dN/d\log M$ is shown in the left panel of figure 4.3. It is customary to model this function at $M \gtrsim 0.4\,M_\odot$ with a power-law (see Fig. 4.2 for a functional form covering all masses). At the end of the formation process the distribution reflects the initial mass distribution from the point of stellar evolution and is referred to as the Salpeter *initial mass function* or IMF

$$\xi(\log m) = \frac{\mathrm{d}N}{\mathrm{d}\log m} \propto m^{-\Gamma}, \tag{4.1}$$

specifying the number of stars in the mass range $(\log m, \log m + d\log m)$. Note that the IMF is related to the *initial mass spectrum*

$$\phi(m) = \frac{\mathrm{d}N}{\mathrm{d}m} \propto m^{-\Gamma-1} \propto m^{-\alpha}, \tag{4.2}$$

i.e. the number of stars in the mass range $(m, m + dm)$, as $d(\log m)/dm = 1/(e\,m)$. The power-law index of the mass spectrum $\alpha = \Gamma + 1$.

The final mass distribution of the simulation has a near-flat slope for low-mass stars, which turns into an increasingly steeper slope for more massive stars. The higher-mass distribution is broadly consistent with a $\Gamma \sim 1$ slope, although it could also be fit by a shallower slope for intermediate-mass stars and a steeper slope for high-mass stars.

The first time this distribution was determined empirically, by Salpeter in 1955 (ApJ 121, 161), a slope of $\Gamma = 1.35$ was found. A more recent determination by Kroupa in 2001 (ApJ 322, 231) yields $\Gamma = 1.3$ for $m > 0.5 M_\odot$, changing to 0.3 in the mass range $0.08 - 0.5 M_\odot$ and to $\Gamma = -0.7$ for masses less than 0.08 $M_\odot$. Note that the Kroupa findings suggest a peak in the IMF in the range ~0.1–0.5 $M_\odot$. Indeed, the simulation is roughly consistent with these observations.

At the end of the simulation, 42 percent of the total mass remains in gas, although much of the gas is no longer bound to the cluster, owing to the turbulence. The *star formation efficiency* is thus 58 percent. The cluster contains 494 $M_\odot$ in a 0.25 pc radius, but only 16 percent is in the form of gas. Realize that the fraction of mass in stars is a function which continuously increases with time (during the formation process) and therefore the value depends on when the simulation is halted.

The predicted star formation efficiency is comparable to that observed for young stellar clusters, but is high if one considers the global (giant) molecular cloud star formation efficiency, which is typically a few to 10 percent. Note however that the simulation does not include any background molecular cloud material not directly involved in the cluster formation process. Still, physical effects not treated in the simulation may be identified that, when accounted for, may reduce the efficiency. Two such effects are *feedback* and *magnetic fields*.
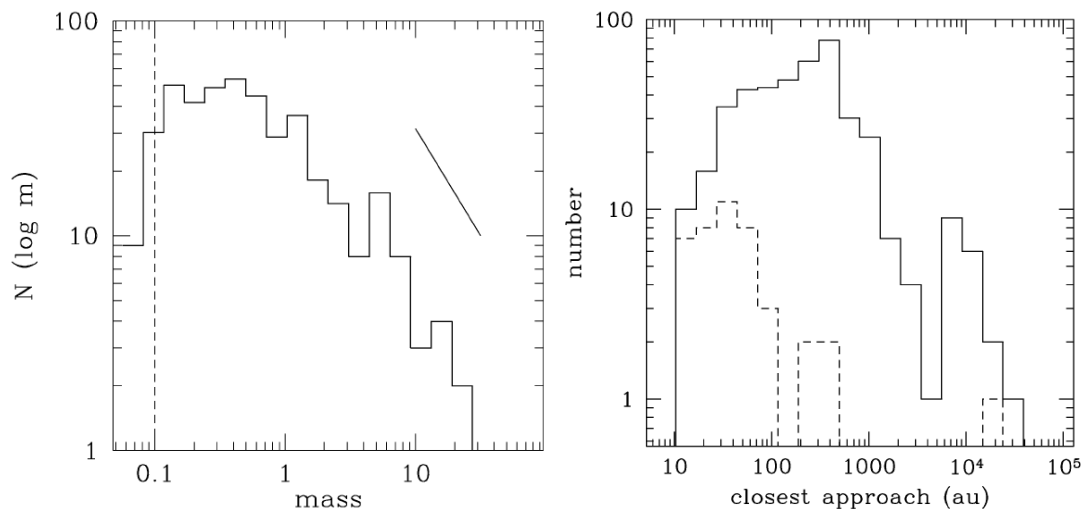
Figure 4.3: *Analysis of the SPH simulation of the collapse of a 1000 $M_\odot$ turbulent molecular cloud forming 418 stars. Left panel: the final mass distribution as a function of the logarithm of the mass. The distribution uses bins in the logarithm of the mass such that a Salpeter IMF has a slope of $\Gamma = -1.35$. The diagonal line notes a $\Gamma = -1$ slope. The highest-mass stars appear to have a steeper distribution, whereas intermediate-mass stars have a shallower one. Stars below roughly a solar mass have an approximately flat distribution. The minimum mass for the simulation is 0.1 $M_\odot$ and is indicated using a dashed line. Right panel: The distribution of minimum closest approaches for all stars (solid line) and stars above 3 $M_\odot$ (dashed line) is plotted as a function of the logarithm of the minimum separation. Gravitational softening is applied for approaches within 160 AU. In these cases the measured values for the closest approaches are upper limits.*

### Feedback and magnetic fields

Forming stars are observed to generate bi-polar outflows. The simple fact that material from the collapsing cores is fed back or re-injected in the molecular cloud reduces the efficiency of star formation. However, the bi-polar outflows are also sources of momentum and energy, re-energizing turbulence in the molecular cloud from within. This adds to the turbulent support of the cloud, which otherwise diminishes once the initial turbulence (that was part of the initial condition of the simulation) decays. This effect too slows down star formation. A further physical effect that must be considered is the feedback on the collapsing cloud resulting from the luminosity of the proto-stars that have formed within it. At first these luminosities are derived mainly from accretion, but later the first (likely most massive) stars add luminosity from contraction and thermo-nuclear burning. The effect at first is to heat the surroundings, increasing the Jeans mass and suppressing further star formation locally, on small scales. Later, radiation pressure exerted by the streaming out of photons may remove gas from the star-forming site altogether. The delay of star formation caused by feedback and the possibility it offers to remove gas altogether will result in a lower star-forming efficiency.

The physical effect of the presence of a magnetic field is to slow down the collapse at large

scales and thus to reduce the rate of star formation.

## 4.3 Interactions

The hierarchical nature of the formation process has many interesting implications for star formation. Sub-clustering means that individual (proto-)stars may enter in regions of higher (proto-)stellar number-density, and experience interaction. The high number-density of stars in each sub-cluster, results in closer and stronger stellar interactions than would otherwise occur. Such stellar interactions can:

- ○ Harden binaries, which may help explain how the closest such systems are formed.

- ○ Truncate circumstellar disks, decreasing their masses and thus their lifetimes.

- ○ Trigger fragmentation in the disc, possibly initiating the formation of one or more stellar or planetary companions.

- ○ Possibly even liberate a population of planets from their parent stars, if planets form quickly.

- ○ Finally, in case the maximum number-density of stars is sufficiently high ($10^7$ to $10^8$ stars $\text{pc}^{-3}$) stellar mergers may play a role in forming the most massive stars.

The right panel of figure 4.3 plots the distribution of closest approaches for each of the 418 stars formed in the simulation. This distribution is calculated, for each star, as the minimum distance to any other passing star at any time during the evolution, and extends from 10 to $> 10^4$ AU. The small peak at large separations indicates the few stars that form in relative isolation in the molecular cloud, and never enter into a (sub-)cluster.

Nearly half of the stars in the simulation have interactions within the 160-AU resolution limit, where gravitational forces start to be smoothed (for numerical reasons; see above). Below 160 AU stars may therefore experience actual closest approaches that are smaller than given in the figure. We see that approximately one-third of the stars have encountered another star within 100 AU, sufficiently close to perturb the circumstellar disk or a binary system. A close passage of a star is likely to truncate a circumstellar disk down to $\sim$ one-third of the minimum separation between stars (Hall et al. 1996), thus limiting their mass reservoirs and lifetimes – although subsequent accretion may replenish the discs.

Figure 4.3 also plots the corresponding distribution of closest approaches for stars more massive than 3 $M_\odot$. The great majority of these massive stars have had a close interaction within 100 AU. Such should then lead to a truncation of the disk to 30 AU or less. This suggests that many of the stars found in young stellar clusters – and virtually all the massive stars – should have small (less than 30 AU) or non-existent disks.

## 4.4 Monolithic Collapse versus Competitive Accretion

The perhaps most striking feature these modern SPH calculations show relates to the mechanism for the formation of massive stars. There has been an ongoing debate regarding this mechanism, which involves primarily the initial condition for the formation. In the classical view, the basic premise is that a single molecular cloud core of $\sim 0.1\,\mathrm{pc}$ would result in the formation of a single, or perhaps a binary, massive star, provided that the mass of the core is high enough. The cloud core has presumably formed by condensation from a much larger unit, a molecular cloud clump with a mass of thousands of $M_\odot$ (or more), but once the core has formed, the resulting collapse is little influenced by the remainder of the material in the clump. Thus low-mass stars would form from individual low-mass cores, and high-mass stars from high-mass cores. The entire mass of the core would not necessarily go into the final star; there is an efficiency factor involved as a result primarily of outflows during the formation process. This model is often referred to as *monolithic collapse*.

The second model, *competitive accretion*, is based on the premise that star formation is controlled not by collapse of individual cores, but by the overall collapse of a much larger region, containing initially gas with several thousand solar masses (or more). The fragmentation into a cluster of stars, containing high-mass and low-mass members, takes place after the overall collapse has started. Individual fragments form at low mass and compete for the accretion of the remaining gas; also there can be interactions among the various fragments. One might expect a massive star to form in the center, since as the cloud collapses and develops a gravitational potential well, a considerable amount of gas can be funnelled toward the central region. In fact massive stars are found to be preferentially located near the centers of clusters, but it is not entirely clear that they formed there. They could have settled to the center after formation as a result of dynamical interactions with other stars. In this picture, then, massive stars must form in clusters, as is generally observed. The computation we have been discussing in this chapter is an example of the competitive accretion model.

Monolythic collapse and competitive accretion imply fundamentally different origins of the IMF of stars. In monolythic collapse, the IMF results from the mass function of cores. In competitive accretion, the IMF results from 'winners' and 'losers' competing for mass from a gas reservoir.

### Closing remark

Although many aspects of the star formation problem still remain unsolved, the introduction of the theory and simulations involving supersonic turbulence have resulted in progress. The observational results over a wide range of scales show that the velocity dispersion scales with the square root of the size of the region; this fact is a strong argument for the existence of supersonic turbulence. The magnetic energy in molecular clouds is known to be comparable to the turbulent kinetic energy. Thus the combination of turbulence and magnetic fields in
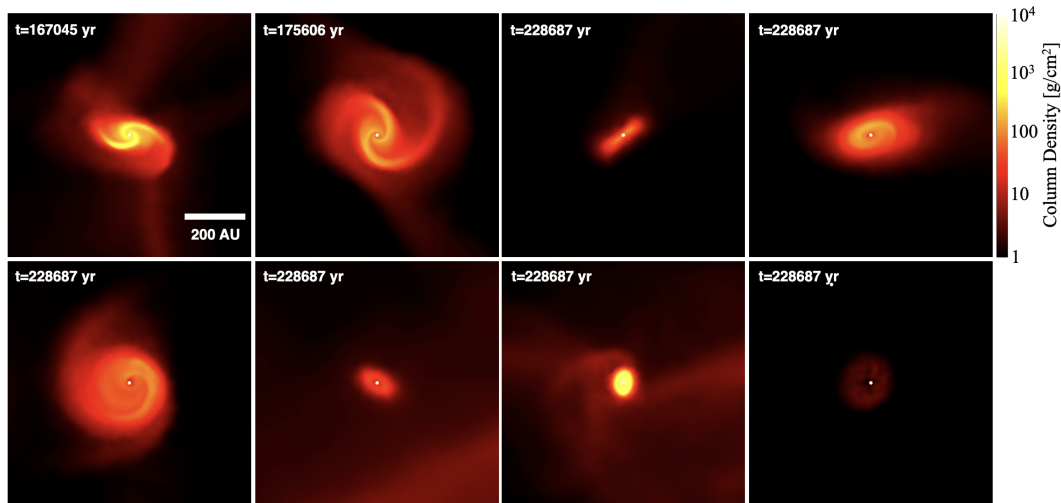
Figure 4.4: *Examples of eight of the circumstellar disks around single protostars from the computation by Bate (2018). The first two cases demonstrate gravitational instabilities in young massive disks soon after they have formed. The remaining six panels show discs at the end of the simulation; they display a wide variety of radii and masses. Sink particles are plotted as white filled circles that have radii 10 times larger than the actual sink particle accretion radius.*

numerical simulations is necessary, although difficult. Eventually, inclusion of even more physical effects in such simulations, along with simultaneous treatment of a wide range of scales, from the size of a whole molecular cloud (parsecs or more) down to a few solar radii, should result in further progress on this key problem.

## 4.5 The diversity and statistical properties of proto-stellar disks

We briefly mention important assumptions in the numerical simulations of Bate (2018, MN-RAS 475, 5618). The molecular cloud contains 500 $M_\odot$ with a diameter of 0.808 pc, giving a mean density of $1.2 \times 10^{-19}$ and an initial free-fall time of the cloud of $\tau_{ff} = 1.90 \times 10^5$ yr. The initial temperature was 10.3 K. An turbulent velocity field was imposed on the medium such that the kinetic energy of the turbulence was equal to the magnitude of the gravitation potential energy of the cloud; this implies a Mach number of 13.7. The calculation used $3.5 \times 10^7$ SPH particles; so each SPH particles contains a mass of $1/70000 \, M_\odot$. Note that this is a factor of 70 more than the Bonnell et al. (2003) simulation discussed at the start of this chapter. Sink particles are introduced to describe the protostar when the density in the second-core collapse phase exceeds $10^{-5} \, \mathrm{gr\,cm}^{-3}$. This density is just two orders of magnitude before the stellar core begins to form (at density $\sim 10^{-3} \, \mathrm{gr\,cm}^{-3}$). A sink particle is formed by replacing the SPH gas particles contained within an accretion radius of 0.5 AU of the densest gas particle in a region. Any gas that later falls within this radius is accreted by the point mass if it is
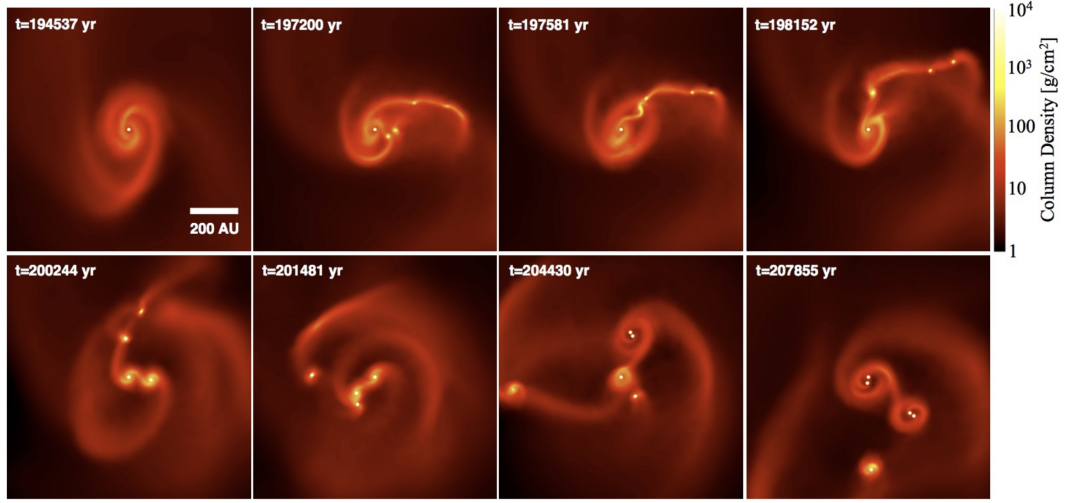
Figure 4.5: *Time sequence showing the fragmentation of a massive disc from the computation by Bate (2018). In panels 2-4 three potential fragments merge into a single object before it collapses to a stellar core. Two more fragments at the top right of the 4th panel eventually collapse to stellar cores and these pair up with the two previously mentioned cores to produce a quadruple system consisting of two pairs. The 5th sink particle (visible in the last two panels) is eventually ejected from the system. Sink particles are plotted as white filled circles that have radii 10 times larger than the actual sink particle accretion radius.*

bound and its specific angular momentum is less than that required to form a circular orbit at 0.5 AU from the sink particle. Thus, gaseous disks, in principle, can be spatially resolved if they have radii $\gtrsim$ 1 AU. However, due to technical constraints posed by the SPH resolution length scaling with density, discs with radii $\lesssim$ 10 AU are not usually resolved in the calculation. Generally, the properties of more massive disks are more reliable than those of low-mass disks. Discs modelled by $\gtrsim$ 2000 SPH particles (i.e., discs with masses $\gtrsim$ 0.03 $M_\odot$) should be well modelled. The sink particles do not contribute radiative feedback. Magnetic fields are not considered.

The cloud was evolved to $t = 1.20\,\tau_{\mathrm{ff}}$ (228 300 yr), by which time 88.2 $M_\odot$ of gas (17.6%) had produced 183 protostars, with a mean mass of 0.48 $M_\odot$. These consisted of 84 single protostars and 40 multiple systems (28 binaries; 5 triples; 7 quadruples). A large population of discs was produces, showing diverse properties. Many of the discs undergo dramatic dynamical evolution during the calculation. This is best appreciated by watching the animation produced by the author, showing a mosaic of 183 animations, each of which displays a region with dimensions $400 \times 400$ AU centred on one of the protostars (sink particles) that is produced during the simulation.

Figure 4.4 shows snapshots of eight discs around single protostars. Many single protostars have large ratios of disc mass to stellar mass soon after they form. Consequently, these discs

display strong spiral arms because they are graviationally unstable (see Sect. 9.5). Some of these fragment, but others are stable enough to avoid fragmentation and transport mass and angular momentum rapidly by gravitational torques from the spiral arms (Laughlin & Bodenheimer 1994). The first two panels give examples of such relatively massive disks. In the first panel the protostar had a mass of $0.20\,M_{\odot}$ and its disc mass was $0.25\,M_{\odot}$. In the second panel the protostar mass was $0.23\,M_{\odot}$ and its disc mass was $0.27\,M_{\odot}$. The remaining six panels show discs at the end of the calculation. The protostars have masses of 0.18, 0.26, 0.29, 0.11, 0.50, and $0.17\,M_{\odot}$, respectively, while their discs have masses of 0.03, 0.25, 0.33, 0.02, 0.38, and $0.004\,M_{\odot}$, respectively. The disc radii are approximately 60, 100, 100, 50, 30, 70 AU in radius, respectively, where in each case this is the radius containing 63.2% of the disc mass.

At various times in the simulation there are more than four dozen single protostars with resolved disks. Not all of these remain single to the end of the simulations. Of those that do, the discs aren't stationary structures. They evolve in time, due to the transport of mass and momentum via gravitational torques, fragmentation, viscous evolution of the disc, and continued mass accretion from the molecular cloud. The discs may also accrete gas or suffer from ram-pressure stripping as they pass through dense cloud material. Star-disc interactions can also strip away or truncate discs, and/or energy loss during a star-disc interaction can produce binaries or high-order multiple systems from protostars that were previously unbound.

The simulation also produces multiple systems (e.g., binaries). Figure 4.5 shows a time sequence of the evolution of a massive disc surrounding a protostar. In the first panel, the proto-stellar mass is $0.07\,M_{\odot}$ while the disk mass is $0.17\,M_{\odot}$. The gravitationally unstable disc has strong spiral arms. In the second panel, four fragments are forming, but in the third and fourth panels three of these merge into a single object, while two further fragments form in the outer part of the largest arm. The fragment resulting from the triple merger does then undergo the second collapse phase and is replaced by a sink particle producing a binary. The two outer fragments also collapse and are replaced by sink particles. The end result is a hierarchical quadruple system. In the meantime a further protostar has formed from the largest spiral arm, resulting in a pentuple system. However, it is eventually ejected from the system.

From these examples, it is clear that the properties and time-evolution of circumstellar disks is complex. Without going into more detail, we briefly mention:

○ Many protostars have their discs eroded or truncated either by ram pressure stripping as they move through dense molecular cloud material, or when they have dynamical encounters with other protostars.

○ Sometimes, discs are eroded and then new discs are being accreted from the molecular cloud.

○ Star-disc encounters are very common. These are frequently involved in forming binary systems (32 cases) or higher-order multiple systems (at least 14 cases). After producing bound systems from two unbound protostars, there is usually rapid decay of the orbital
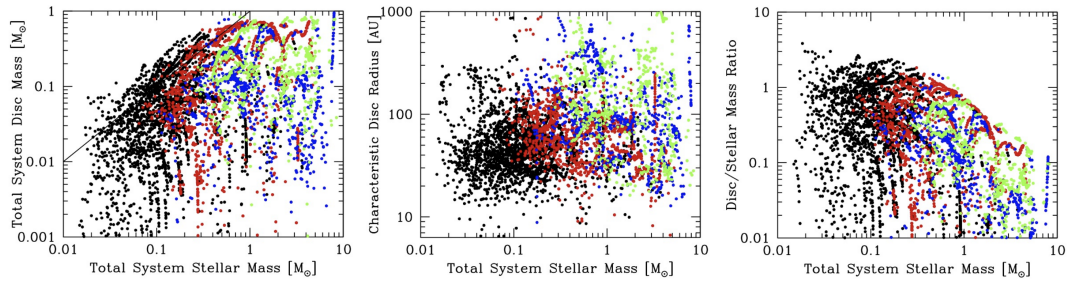
Figure 4.6: *The total disc masses (left), characteristic disc radii (centre), and the ratio of total disc mass to total stellar mass (right) of the proto-stellar systems versus their total stellar mass from the computation by Bate (2018). Each dot represents an instance of disc(s) of a particular system, which may be a single star or a bound multiple proto-stellar system. A single system may be represented by many dots that give the state of the system at different times. The colours denote the order of the system: single (black), binary (red), triple (blue), or quadruple (green). Disc masses tend to be greater for more massive proto-stellar systems, but the scatter is large. Discs around very low mass stars ($< 0.1\,M_\odot$) are noticeably smaller than those around more massive systems. For more massive systems, the typical disc size does not depend strongly on the total stellar mass, but the largest discs tend to be found around some of the most massive systems, and these tend to be multiple systems. The typical ratios of the disc to stellar mass tend to be highest for systems with intermediate masses ($0.1 - 0.3\,M_\odot$) and they tend to decline strongly with increasing stellar mass for stars with mass $> 0.3\,M_\odot$.*

      separation and eccentricity as the binary transfers angular momentum and energy to the dissipative gas, often producing a circum-binary disc.

○ Discs in all the systems in the simulation include circumstellar, circum-binary, and/or circum-multiple discs. One higher-order multiple system may show all these types of discs. These may all be oriented differently relative to one another.

○ After a proto-stellar system has formed, its disc can continue to accrete gas from the cloud. Since the cloud is turbulent, the orientation of the angular momentum of this additional gas may be different form that that formed the initial disk. This may even produce discs in which the inner and outer parts have different orientations.

Also worth mentioning is:

○ Sink particles were permitted to merge if they passed within 0.01 AU ($\sim 2\,R_\odot$) of each other. However, no mergers occurred during the simulation. This is not to say that perhaps later on in the formation process such mergers may occur.

Figure 4.6 shows how the total disc mass, characteristic disc radius, and total disc mass to stellar mass depend on total stellar mass. If the star is, e.g., a binary and has a circum-binary disc that total stellar mass is the sum of the mass of the two stars. A single proto-stellar system (a single stars, binary, triple, or quadruple) may appear multiple times in the plots, reflecting

different moments in its evolution. A large diversity of possibilities is apparent, with some weak trends: (1) Discs tend to be more massive for more massive systems; (2) Discs around very low mass stars are, on average, notably smaller than those around more massive systems; (3) Disc sizes reach up to $1000\,\mathrm{AU}$, with the largest discs found mostly around the most massive systems, and these tend to be multiple systems; (4) The typical ratio of disc mass to total stellar mass seems to be at about $0.4\,M_\odot$ proto-stellar systems and tends to decline strongly with increasing proto-stellar system mass.

By the end of the simulation, part of the young proto-star population is left with little or no disc material. Though partly this may be due to the limited numerical resolution, most of the destruction of discs is due to dynamical interactions between protostars, with ram-pressure stripping also having a role. There is observational evidence that this is also the case in nature. Typical fractions range from $\sim 50 - 85\,\%$. The highest disc fractions ($\sim 85\,\%$) have been found in the Orion Nebula Cluster NGC 2024, NGC 1333, and NGC 2068/71.

## Closing remark

The important thing to take away from these simulations of proto-stellar disc properties is that the geometrical structure of the discs is often complex and can be highly time-dependent. Later on in these lecture notes, we will present analytical models of these discs that allow to discuss their properties in some more detail. Of course, we can only do this after making assumptions, such that they are axial symmetric and hydrostatic. Keep in mind that these are very strong assumptions.

## SMOOTHED PARTICLE HYDRODYNAMICS WITH SINK PARTICLES

Many simulations of star formation in turbulent interstellar clouds are conducted using a Lagrangian method to solve the equations of hydrodynamics. This is called smoothed particle hydrodynamics (SPH). It is a 3D method that assumes no symmetries and uses no imposed grids. Monaghan (2005) gives a recent review of the method and its applications (see also Monaghan 1992). With SPH, the gas is represented by a number of sampling points ("particles"), each associated with a certain gas mass. The mass of each particle is distributed in space according to a spread function so that density and other hydrodynamic quantities are continuous in space. The particles are allowed to move in the computational domain under the laws of self-gravitating hydrodynamics (which are evaluated at the particle positions). The system is evolved in time using a standard integration routine.

The timestep with which the simulations evolve is chosen such that all changes to the system's state are examined at the appropriate speed. For example, during the collapse phase, when infalling particles approach each other, the timestep gets progressively shorter. This may potentially bring the simulations to an early halt as the timestep converges to zero. This problem is overcome with the introduction of star (sink) particles of finite radii that each replace groups of dense neighboring gas particles above a certain density threshold (see Bate, Bonnell & Price 1995). The sink particles are then evolved as collisionless matter that can accrete from the lower density gas whenever gas particles approach within the sink radius. The final state of the system in such simulations involves groups of sink particles in spatial distributions resembling observed young clusters. However, all information on scales smaller than the sink radius is lost; such sink particles cannot later become sources of outflows, unless specifically instructed to do so. Note that sink particles are also usefully employed in Eulerian collapse calculations, as shown by Krumholz, McKee & Klein (2004).

In order for fragmentation to be realistically modelled, the numerical resolution of the SPH simulations must be such that they obey the Jeans condition. This states that "the local Jeans mass must always remain resolved," i.e., it must contain more than a certain number of particles (Bate & Burkert 1997; Truelove et al. 1997; Kitsionas & Whitworth 2002; Martel, Evans & Shapiro 2006). The Jeans mass decreases as the density increases up to the point at which the gas stops being approximately isothermal and begins to heat up adiabatically. From there on, the Jeans mass starts increasing again with density. The density at which this occurs defines the minimum Jeans mass that the simulation has to resolve and, thus, the minimum number of particles required. Present-day computer capabilities allow the use of several million SPH particles per simulation, a number that is nominally adequate for resolving the density at which adiabatic heating switches on. Using parallel computing facilities presently available, such computations take several months for each evolutionary model.

Figure 4.7: *Discussion on SPH from Zinnecker & Yorke, 2007, ARA&A 45, 481 (page 497).*

### Exercise 4.1

a) All stars initially more massive than 8 $M_\odot$ explode as supernova. Assume that all stars form in isolation (so no binaries or higher order multiples form) and that the initial mass function is according to Salpeter. If stars are formed in the mass range from 0.1 $M_\odot$ to 300 $M_\odot$, which fraction of stars that form in a co-eval burst leave supernova?

b) All stars more massive than 15 $M_\odot$ are born as O-type stars. It is these stars that produces hydrogen ionizing photons. Which fraction of stars that form in a co-eval burst have this spectral type? Use the same IMF assumptions as in (a).

c) All stars less massive than 0.5 $M_\odot$ are born as M-type stars. The habitable zone (or water zone) is for these stars so close to the stellar surface that potential terrestrial planets in this zone are likely tidally locked. If so, the night side of the planet will be extremely cold. This is expected lead to a freeze-out of a possible planetary atmosphere, making such world unsuited to harbor life. Which fraction of stars that form in a co-eval burst have this spectral type? Use the same IMF assumptions as in (a).

### Exercise 4.2

We consider motions in a spherical stellar cluster with a constant stellar density $\rho_\star$. The total cluster mass in the cluster is $M_c = 10^6 M_\odot$ and the cluster radius is $R_c = 1\,\mathrm{pc}$.

a) Give the gravitational acceleration in the cluster as a function of distance from the center.

b) Give and solve the equation of motion for a particle (for instance a star) released at a distance $r_\circ$ from the cluster center with a velocity $v = 0$. Such a particle will fall through the center, so we can consider the orbit as one-dimensional, i.e. to depend on $r$ only. In solving this problem, use the ansatz $r(t) = r_\circ \cos \omega t$.

c) The dynamical time $t_{\mathrm{dyn}}$ in a cluster is how long it takes a particle to complete a single orbit. Give the expression for $t_{\mathrm{dyn}}$ and compute its numerical value.

d) Compute the velocity for the particle as a function of time. What is the maximum velocity $v_{\mathrm{max}}$, and where is it reached?

e) Assume that all stars in the cluster have the same total energy. How, then, must $r_\circ$ depend on the mass of the star? Where in the cluster would you expect to find the massive stars, and where will the low-mass stars be?

f) We assume all stars to have a mass of $M_\star = 1 M_\odot$ and a radius of $R_\star = 10 R_\odot$. Give an expression for the collision time in the center of the cluster, using the geometrical cross section of a star, the stellar density and the maximum speed $v_{\mathrm{max}}$.

g) If the surface escape velocity $v_{esc}$ of the stars in the cluster is larger than the typical speed $v$ of the stars, then the cross section for collisions is enhanced by a factor

$$f_{Saf} = 1 + \left(\frac{v_{esc}}{v}\right)^2.$$
(4.3)

This is the gravitational focussing or Safronov factor, which will be discussed in Sect. (12.1). What is the collision time in the cluster center if you take this factor into account?

# 5

# Cloud equilibrium and stability

In chapters 2 and 3 we have discussed aspects of the basic structure of molecular clouds. We now turn to small cloud cores and study the density structure of these cores and consider their stability, and the conditions that must occur for collapse to proceed.

An example of a relatively isolated self-gravitating core is Barnard 68 in the $\rho$ Ophiuchi cloud, already shown in figure 2.4 and again in figure 5.1. $\rho$ Ophiuchi is a rather small star formation region, hence the cores in this region are relatively isolated. Such relatively isolated cores are named *Bok globules*, in honor of the Dutch astronomer Bart Jan Bok who first recognized, in the 1940s, their potential role in stellar birth. Larger and denser star forming regions such as the Orion Molecular Cloud also contain such cores, however comparative studies show that in these regions the cores are systematically warmer, by about 5 K. These objects also have $NH_3$ line widths two to three times broader, indicating a higher degree of turbulent support. Moreover, the elevated gas temperatures close to Orion's hot massive stars promote a qualitatively distinct chemical reaction network. Finally, the typical Orion dense core has three times the mass and twice the diameter of its counterpart in $\rho$ Ophiuchi or Taurus-Auriga.

The size of Barnard 68 is $0.074 \times 0.047$ pc from continuum measurements at $850\,\mu$m, its temperature as derived from the spectral energy distribution is $13 \pm 2$ K and its visual extinction $A_V$ is 11 mag. The average particle density in the direction of the center of the cloud is estimated to be $5 \times 10^4$ cm$^{-3}$ (see Kirk et al. 2005, MNRAS 360, 1506 for details). Alves et al (2001, Nature 409, 159) found the density distribution of this core to be super-critical and hence postulate that it is at the verge of collapse and should eventually form a low mass star.

In the following, we discuss the density structure of such cores and the role of magnetic fields and turbulence in their stability.
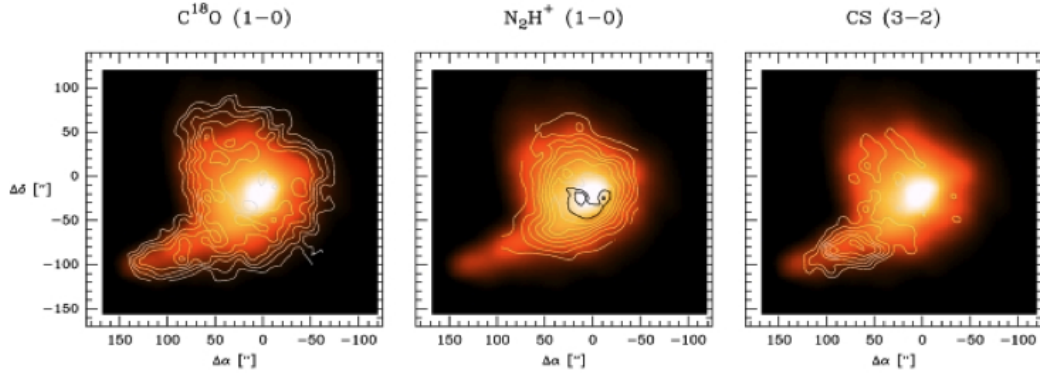
Figure 5.1: *Images of the dark cloud Barnard 68 (see also figure 2.4). The yellow color map shows the continuum emission from dust, a good measure of the mass distribution. Over-plotted in the left panel are contours of $C^{18}O$ 1-0 line emission, which traces the CO gas column density. The CO column density increases with the dust emission, except towards the center where it remains constant although the dust column density keeps increasing. This is because the CO molecules condense out onto the grains. $N_2H^+$ tracks the dust better towards the center, whereas CS disappears from the gas phase already in the outer regions.*

## 5.1  Isothermal, self-gravitating spheres

As the most simple and idealized case, we now turn to clouds which are spherical, isothermal, and where gravity is only balanced by pressure forces. The assumption of spherical symmetry is a rather crude one as can be seen by looking at figure 5.1 or 2.4. The assumption of an isothermal state simplifies the calculations because in this case pressure gradients are solely due to density gradients, and a simple, scale-free formulation can be found. It is clear that this is an over-simplification. In reality, magnetic support and magneto-hydrodynamic waves play a major role in determining the core structure, at least in larger cloud complexes. However, the basic concepts of stability and collapse can be captured here.

We first write down the basic equations for hydrostatic equilibrium for such systems. These are

$$\frac{\mathrm{d}p}{\mathrm{d}r} = -\frac{G\,M\,\rho}{r^2} = -\frac{\mathrm{d}\Phi}{\mathrm{d}r}\,\rho, \quad \text{and} \tag{5.1}$$

$$p = \frac{\rho\,k\,T}{\mu m_\mathrm{H}} = \rho\,a^2, \tag{5.2}$$

where the first equation is that for hydrostatic equilibrium (see equation 3.23) and the second is the equation of state 3.2 for an ideal gas. $a$ is the isothermal sound speed in the gas at temperature $T$. We complement these equations with an equation for the gravitational potential $\Phi$, i.e. Poisson's equation

$$\nabla^2\,\Phi = 4\pi G\,\rho, \tag{5.3}$$

with the right-hand side referring to the density of the cloud itself. In spherical symmetry

$$\nabla^2 \Phi = \Delta \Phi = \nabla \cdot \nabla \Phi = \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial \Phi}{\partial r} \right). \tag{5.4}$$

With $\Phi = -GM/r$, one can reformulate this to

$$\frac{\partial M}{\partial r} = 4\pi \, r^2 \, \rho. \tag{5.5}$$

In spherical symmetry we may replace the partial derivative $\partial/\partial r$ by $d/dr$. To integrate this set of equations one has to specify boundary conditions. We can specify the density at the core of the cloud, $\rho(0)$, and directly integrate the system from the inside out, knowing that $M(0) = 0$. The integration can be terminated at a location $r_{\text{out}}$ where a specific outer pressure $p = \rho(r_{\text{out}}) \, a^2$ is attained. Equation (5.5) allows to calculate the mass $M(r_{\text{out}})$ of the system.

If we want to construct a cloud with a given mass and a given external pressure, then we have to construct a series of models with different values for $\rho(0)$. Figure 5.2 shows a number of such integrations for different central density. It shows that all solutions asymptotically merge into the same power-law in the outer regions. The asymptotic solution for infinite central density is the so-called *singular isothermal sphere*

$$\rho(r) = \frac{a^2}{2\pi G \, r^2}, \tag{5.6}$$

showing that far from the center the density decreases as $1/r^2$. It is straightforward to show that for this solution

$$M(r) = \frac{2a^2}{G} \, r, \tag{5.7}$$

where $M(r)$ is the mass up to radius $r$.

Figure 5.2 shows that different solutions of these equations look similar, as if they are scaled versions of each other. This indicates that it is useful to try a scale-free formulation of the problem.

## Scale-free formulation

We can use equations (5.1) and (5.2) to express hydrostatic equilibrium in terms of the gas density alone. This yields

$$\frac{1}{\rho} \frac{d\rho}{dr} = \frac{d \ln \rho}{dr} = -\frac{1}{a^2} \frac{d\Phi}{dr}. \tag{5.8}$$

Equation (5.8) implies that the sum $(\ln \rho + \Phi/a^2)$ is independent of $r$, so we are going to explore solutions in the form

$$\rho(r) = \rho_c \, \exp\left( \frac{-\Phi}{a^2} \right) = \rho_c \, \exp\left( -\Psi \right), \tag{5.9}$$
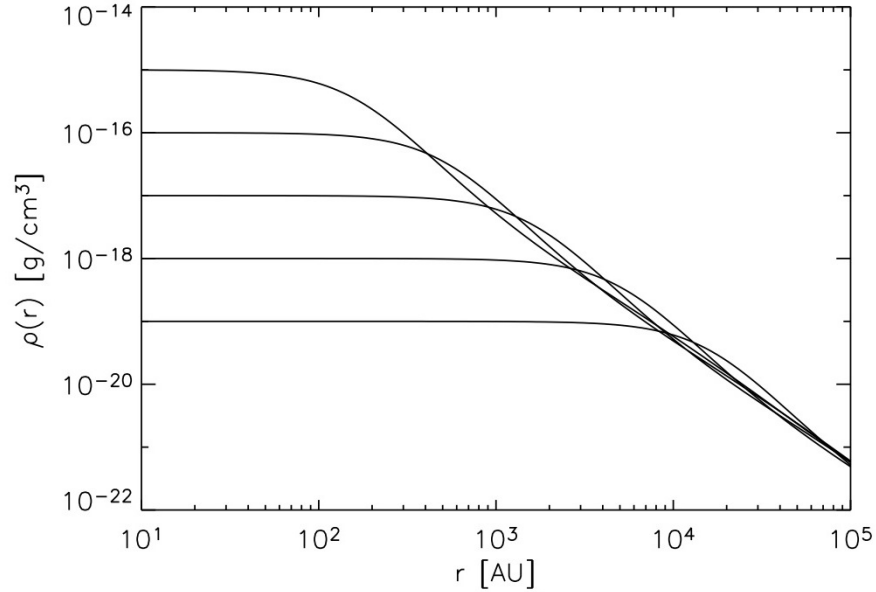
Figure 5.2: *Solutions for the density profile of an isothermal sphere in hydrostatic equilibrium, assuming different values for the central density.*

where $\rho_c = \rho(0)$, the central density of the cloud, and where we have introduced a dimensionless potential

$$\Psi(r) = \frac{\Phi}{a^2}. \tag{5.10}$$

Using $\Psi$ we may rewrite the equation of hydrostatic equilibrium to

$$\frac{d\Psi}{dr} = -\frac{1}{\rho\, a^2} \frac{dp}{dr}. \tag{5.11}$$

We now introduce a dimensionless distance

$$\xi = \left(\frac{4\pi G\, \rho_c}{a^2}\right)^{1/2} r. \tag{5.12}$$

Manipulating equations (5.3) through (5.5), using (5.10) and (5.11) yields

$$\frac{1}{4\pi\, r^2} \frac{dM}{dr} = -\frac{1}{4\pi\, r^2} \frac{d}{dr}\left(\frac{r^2}{G\rho} \frac{dp}{dr}\right) = \rho. \tag{5.13}$$

Substituting subsequently $dp/dr$ from equation (5.11) and then $\xi$ we obtain

$$\frac{1}{4\pi\, r^2} \frac{d}{dr}\left(\frac{r^2\, \rho\, a^2}{G\rho} \frac{d\Psi}{dr}\right) = \left(\frac{4\pi G\, \rho_c}{a^2}\right) \frac{1}{4\pi\, \xi^2} \frac{d}{d\xi}\left(\xi^2 \frac{a^2}{G} \frac{d\Psi}{d\xi}\right) = \rho = \rho_c \exp\left(-\Psi\right), \tag{5.14}$$

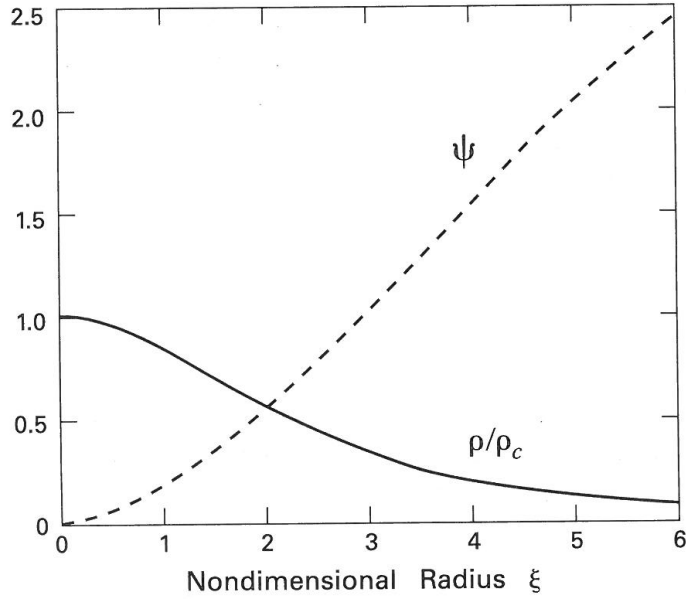Figure 5.3: *Non-dimensional gravitational potential (dashed curve) and density (solid curve) in a spherical isothermal cloud, as a function of the non-dimensional radius. From: The Formation of Stars, Stahler & Palla, 2011.*

and finally

$$\frac{1}{\xi^2}\frac{d}{d\xi}\left(\xi^2\frac{d\Psi}{d\xi}\right) = \exp(-\Psi). \tag{5.15}$$

This is the *isothermal Lane-Emden equation*. The boundary conditions of this equation at the center of the cloud, i.e. at $\xi = 0$, can be specified realizing that at that point both the gravitational potential $\Phi = a^2\,\Psi$ and acceleration $\mathbf{g} = -\nabla\Phi = -a^2\,\nabla\Psi$ are zero. Summarizing:

$$\Psi(0) = 0 \qquad \text{and} \qquad \left.\frac{d\Psi}{d\xi}\right|_{\xi=0} = 0. \tag{5.16}$$

The Lane-Emden equation can be integrated numerically to find $\Psi$ as a function of $\xi$. The dashed curve in figure 5.3 displays $\Psi(\xi)$. Of greater interest is the ratio $\rho/\rho_c$, which, from equation (5.9), is given by $\exp(-\Psi)$. At large distances ($\xi \gg 1$), $\rho/\rho_c$ asymptotically approaches $2/\xi^2$, hence $\Psi = \ln(\xi^2/2)$.

The pressure in such a cloud must fall monotonically from the center to the outside, in order to offset the weight of each added layer of material. In an isothermal cloud this means that the density will fall monotonically. In principle, we can follow this integration to arbitrarily large radii. However, in any actual cloud, the pressure does not fall to zero, but to some value $p_o$ characterizing the external medium. For given $a$ we can immediately derive the density $\rho_o$ at the outer cloud boundary, using equation of state (5.2).

How does figure 5.3 tell us the properties of the cloud, given these constraints? To do this we suppose a certain density contrast from center to edge, $\rho_c/\rho_o$. Using the inverse of this ratio, we may read off from the figure the corresponding non-dimensional radius of the outer cloud boundary $\xi_o$. The corresponding radius $r_o$ follows simply by inversion of equation (5.12). Essentially, figure 5.3 describes a sequence of models for $\rho(r)$, out to $r_o$, conveniently parametrized by the density contrast $\rho_c/\rho_o$.

Let us now consider the mass $M(r_o)$. Integration over spherical shells gives

$$M(r_o) = \int_0^{r_o} 4\pi\, r^2\, \rho(r)\, dr = 4\pi\, \rho_c \left( \frac{a^2}{4\pi G\, \rho_c} \right)^{3/2} \int_0^{\xi_o} e^{-\Psi}\, \xi^2\, d\xi. \tag{5.17}$$

The integral is easily evaluated using equation (5.15). We find

$$\int_0^{\xi_o} e^{-\Psi}\, \xi^2\, d\xi = \int_0^{\xi_o} \frac{d}{d\xi} \left( \xi^2 \frac{d\Psi}{d\xi} \right) d\xi = \left[ \xi^2 \frac{d\Psi}{d\xi} \right]_0^{\xi_o} = \xi_o^2\, \frac{d\Psi}{d\xi}\bigg|_{\xi_o} - 0 \cdot 0, \tag{5.18}$$

because at $\xi = 0$ also $d\Psi/d\xi = 0$ (see equation 5.16).

If we define a dimensionless mass

$$m(\xi_o) = \frac{p_o^{1/2}\, G^{3/2}\, M}{a^4}, \tag{5.19}$$

conveniently using the assumed $p_o$ and $a$, our final result is that

$$m(\xi_o) = \left( 4\pi\, \frac{\rho_c}{\rho_o} \right)^{-1/2} \xi_o^2\, \frac{d\Psi}{d\xi}\bigg|_{\xi_o} \tag{5.20}$$

The value of $\xi_o$ is known for each density contrast $\rho_c/\rho_o$. Thus, the last factor on the right-hand side of (5.20) may be extracted from the $\Psi(\xi)$ curve in figure 5.3.

The dimensionless mass obtained in this manner is plotted as a function of the density contrast in figure 5.4. If the density were constant throughout, so $\rho_c = \rho_o$, it would follow that $\xi_o = 0$, implying that $m = 0$. With increasing density contrast, $m$ first rises to a maximum value of $m_{\text{max}} = 1.18$ attained at $\rho_c/\rho_o = 14.1$. The mass then drops to a minimum of $m_{\text{min}} = 0.696$, eventually approaching in an oscillatory fashion the asymptotic limit $m_\infty = (2/\pi)^{1/2} = 0.798$. One may verify directly from equation (5.6) that $m_\infty$ represents the non-dimensional mass of the singular isothermal sphere.

## Gravitational stability of the Lane-Emden equation

It turns out that only a limited subset of the full model sequence (in $\rho_c/\rho_o$, for given external pressure $p_o$ and isothermal sound speed $a$, implying a certain cloud radius $r_o$ and cloud mass $M$) is *gravitationally stable*. In all other clouds, an arbitrary small initial perturbation in the structure grows rapidly in time, leading ultimately to collapse.
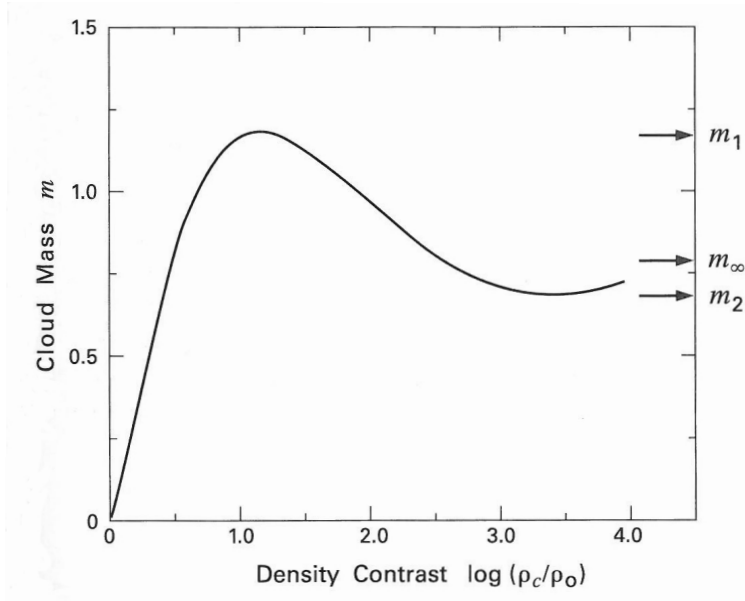
Figure 5.4: *Non-dimensional mass of pressure-bounded, isothermal spheres. The mass is shown as a function of the density contrast from center to edge. From: The Formation of Stars, Stahler & Palla, 2011.*

For a stable cloud, any increase of the external pressure $p_o$ creates both a global compression and a rise of the internal pressure, where the latter acts to re-expand the configuration. Clouds of low density contrast are in this pressure-dominated regime. For clouds with a large $\rho_c/\rho_o$ it is much more difficult for the central regions to expand after application of an enhanced $p_o$. Such clouds are gravity dominated, and will become gravitationally unstable 'when pushed against'. Stability analysis shows that solutions in figure 5.4 with a negative slope (i.e. $dm/d(\rho_c/\rho_o) < 0$) are unstable, so that the maximum $m = 1.18$ corresponds to the maximum density contrast of a stable solution. All clouds with $\rho_c/\rho_o > 14.1$, i.e. those to the right of the first maximum, are gravitationally unstable. The mass corresponding to the critical value is known as the *Bonnor-Ebert mass*:

$$M_{\mathrm{BE}} = \frac{m_{\mathrm{BE}}\, a^4}{p_o^{1/2}\, G^{3/2}} = \frac{m_{\mathrm{BE}}\, a^3}{\rho_o^{1/2}\, G^{3/2}}, \qquad (5.21)$$

where $m_{\mathrm{BE}} = 1.18$.

Extracting the dimensionless distance corresponding to the density contrast $\rho_c/\rho_o = 14.1$ of the Bonnor-Ebert solution from figure 5.3 yields $\xi_{\mathrm{BE}} = 6.5$. The radius of the Bonnor-Ebert sphere is then extracted from equation (5.12) and is equal to

$$r_{\mathrm{BE}} = 1.83 \left(\frac{a^2}{G\,\rho_c}\right)^{1/2} = 0.76 \left(\frac{a^2}{G\,\bar{\rho}}\right)^{1/2} = 0.49 \left(\frac{a^2}{G\,\rho_o}\right)^{1/2}, \qquad (5.22)$$
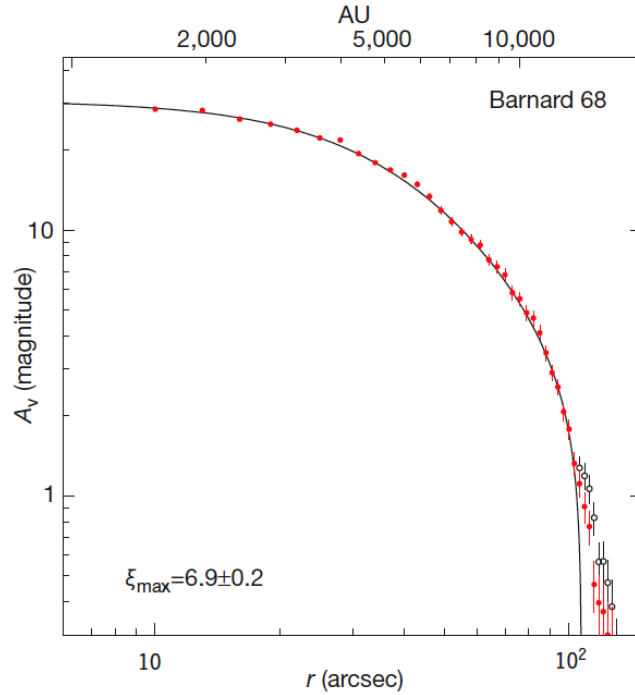
Figure 5.5: *Azimuthally averaged radial dust column density profile of Barnard 68. By convention the dust column density is expressed in terms of magnitudes of visual extinction $A_V$. The solid line represents the best fit of a theoretical Bonnor-Ebert sphere to the data. From: Alves et al. 2001, Nature 409, 159. See also Alves et al. 2001, Messenger 103, 1.*

where the second right-hand side has been computed from the density profile as displayed in figure 5.3. This result is very much comparable to that obtained in the Jeans criterion analysis in section 3.3 as the first right-hand side of equation (3.62) implies that

$$r_J = \left( \frac{15 \, a^2}{4\pi G \, \rho} \right)^{1/2} = 1.09 \left( \frac{a^2}{G \, \rho} \right)^{1/2}, \tag{5.23}$$

where $\rho$ in the Jeans instability analysis is assumed to be a constant density. Likewise, the Bonnor-Ebert mass (equation 5.21) is very much similar to the Jeans mass (equation 3.61).

Figure 5.5 shows a fit of Bonnor-Ebert models to the observed density distribution of Barnard 68. The best solution has a value $\xi_\circ = 6.9 \pm 0.2$ at the outer cloud boundary. For the known distance (125 pc), and temperature (16 K), B68 has a physical radius of 12,500 AU or 0.06 pc, a mass of 2.1 $M_\odot$ and an outer boundary pressure $p_\circ = 2.5 \times 10^{-12}$ Pa. This surface pressure is an order of magnitude higher than that of the general ISM but it is in rough agreement with the pressure inferred for the Loop I superbubble, in which Barnard 68 is embedded. The close correspondence of the observed extinction profile with that predicted for a Bonnor-Ebert sphere strongly suggests that B68 is indeed an isothermal, pressure confined, and self-gravitating cloud. It is also likely to be in a state near hydrostatic equilibrium with thermal pressure pri-

marily supporting the cloud against gravitational contraction. For Barnard 68, $\xi_\circ$ is very near and slightly in excess of the critical radial parameter and the cloud may be only marginally stable and on the verge of collapse.

Ways to speed up collapse of Barnard 68 include increasing the external pressure so that the density contrast in the sphere eventually exceeds the Bonnor-Ebert limit, or (with constant external pressure) to load mass into the sphere until the maximum mass is exceeded.

## 5.2 Rotational support

Rotation is an inherent part of motions in the Milky Way. Not only do all clouds orbit about the center of the galaxy, but turbulence in the interstellar medium is an inherently rotational phenomenon involving the formation of vortices. Indeed, it is observed that clouds and the core therein rotate. The velocity structure of (starless or pre-stellar) cores can be observed with molecular line tracers such as $NH_3$, $N_2H^+$, CS or $C_2S$ (see e.g. Fig. 5.1). Figure 5.6 shows maps of the latter molecule for the starless core Lynd1544 (or L1544). Spatial variations of the location of the line peak measured relative to the local standard of rest velocity $v_{LSR}$ reveal motions of the core. Typical velocity gradients observed in this way range from 0.3 to $10 \, \text{km s}^{-1} \, \text{pc}^{-1}$ (see for example André et al. 2007, A&A 472, 516). Does this mean that rotation plays a role in the gravitational stability of these cores?

To answer this question, we compare the rotational and potential energy of the core. The rotational kinetic energy of a uniform-density sphere with mass $M$ and radius $R$ is

$$E_{\text{rot}} = \frac{1}{2} I \omega^2 = \frac{1}{5} M R^2 \omega^2, \tag{5.24}$$

as for such a sphere the moment of inertia $I = (2/5) M R^2$. Here, $\omega$ is the angular velocity of the cloud. The velocity gradient mentioned above can be interpreted as $\omega \simeq \Delta v_{\text{rot}} / \Delta \ell$, where $\ell = 2R$ is the size of the cloud, if we assume that the cores are solid rotators with their axis perpendicular to the line of sight. The potential energy of a constant density sphere is given by equation (3.55). Hence, the ratio of the two becomes

$$\frac{E_{\text{rot}}}{E_{\text{grav}}} \approx 0.01 \times \left( \frac{\omega}{1 \, \text{km s}^{-1} \, \text{pc}^{-1}} \right)^2 \left( \frac{R}{0.1 \, \text{pc}} \right)^3 \left( \frac{M}{10 \, M_\odot} \right)^{-1}. \tag{5.25}$$

This clearly shows that for the typical observed cases of slowly rotating cores, the centrifugal forces will not dominate over gravitational forces in the initial stage of the collapse.

Numerical simulations of the collapse of isothermal rotating clouds have shown that significant flattening only occurs above $E_{\text{rot}}/E_{\text{grav}} \sim 0.1$. Accounting for rotation does not affect the qualitative shape of figure 5.4, but the peak of the dimensionless mass shifts with increasing rotation energy to higher $m$. This illustrates the basic principle, namely that rotational support (by the centrifugal force) tends to stabilize a core against collapse.
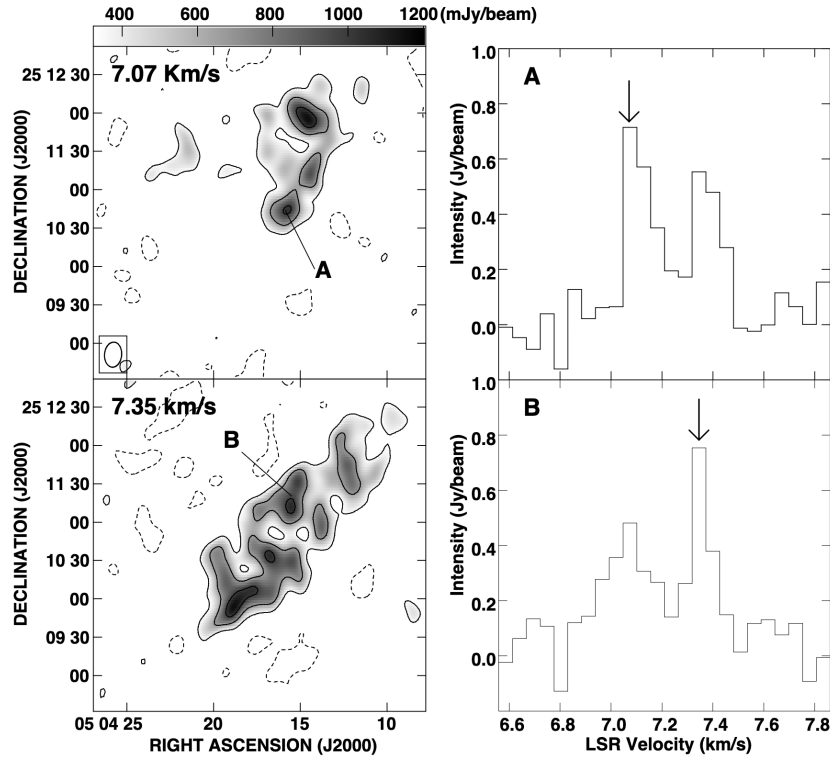
Figure 5.6: *Maps (left panels) of the emission in the $C_2S$ ($J_N = 3_2 - 2_1$) line at 9 mm of the starless core Lynd1544 (or L1544), using the Berkeley-Illinois-Maryland Association array. The maps show an elongated condensation, $0.15 \times 0.045$ pc in size, with stronger emission at the edges. The appearance is consistent with a flattened, ringlike structure viewed at high inclination to the line of sight. The $C_2S$ molecule is likely heavily depleted in the inner part of the core. The position velocity diagram along the major axis can be reproduced by a simple model ring with motions of both infall and rotation. The models suggest comparable velocities for infall and rotation, $\sim$0.1 $\mathrm{km\,s^{-1}}$, in the outermost envelope, at radius 15 000 AU. $C_2S$ line profiles of representative clumps (A and B) are shown in the left panels. A double velocity component is obvious for each line profile. The arrows indicate the corresponding velocity component for each clump. From: Ohashi et al. ApJ 518, L41 (1999).*

Rotation will play a role later as we study the final stages of collapse and especially the formation of disks.

## 5.3 Magnetic support

As magnetic fields are expected to play a role in the support of giant complexes (see section 3.2) it seems prudent to estimate their role in supporting cores against collapse. Along the same line as above, we can compare the magnetic energy to the gravitational energy. The

| Object | Type of Region | Diagnostic | $B_{\parallel}\,(\mu G)$ |
|--------|----------------|------------|--------------------------|
| Ursa Major | Diffuse Cloud | H I | $+10$ |
| L204 | Dark Cloud | H I | $+4$ |
| NGC2024 | GMC Clump | OH | $+87$ |
| B1 | Dense Core | OH | $-27$ |
| S106 | H II Region | OH | $+200$ |
| Sgr A/West | Molecular Disk | H I | $-3000$ |
| W75 N | Maser | OH | $+3000$ |

Table 5.1: *Zeeman measurements of magnetic fields. In all cases, it is the line-of-sight component $B_{\parallel}$ that is actually observed, through Zeeman splitting of either the 21 cm H I line or the hyperfine OH lines near 18 cm. From: Stahler & Palla 2011, The Formation of Stars (Table 9.1 page 257).*

magnetic energy can be derived for a constant large-scale magnetic field $B = |\mathbf{B}|$ as

$$E_{\text{mag}} = \frac{1}{8\pi} \int B^2 \, dV = \frac{1}{6} B^2 \, R^3 \tag{5.26}$$

Comparing this to the gravitational potential results in

$$\frac{E_{\text{mag}}}{E_{\text{grav}}} = \frac{5}{18} \frac{B^2 \, R^4}{GM^2} \approx 11.7 \left( \frac{B}{20\,\mu G} \right)^2 \left( \frac{R}{0.1\,\text{pc}} \right)^4 \left( \frac{M}{10\,M_\odot} \right)^{-2} \tag{5.27}$$

Stahler & Palla present a table with Zeeman measurements of line-of-sight magnetic field strength $B_{\parallel}$ in clouds and cores (see table 5.1). The table entry for the dense core suggests that indeed magnetic forces are important.

So indeed, when a magnetic field is present in the cloud, and when there are sufficient charges in the cloud to keep the magnetic field frozen in the gas, magnetic pressure adds to the stability of the cloud. What will happen to the Jeans mass?

We first write down the virial theorem (equation 3.45) in the presence of a magnetic field. To this end we add the magnetic energy $E_{\text{mag}}$, so that

$$2 \left\langle E_{\text{th}} \right\rangle + \left\langle E_{\text{mag}} \right\rangle + \left\langle E_{\text{grav}} \right\rangle = 0. \tag{5.28}$$

If the magnetic field is the dominant support mechanism (i.e. $E_{\text{mag}} \gg E_{\text{th}}$), then for equilibrium, the magnetic virial theorem reduces to

$$\left\langle E_{\text{mag}} \right\rangle = - \left\langle E_{\text{grav}} \right\rangle \tag{5.29}$$
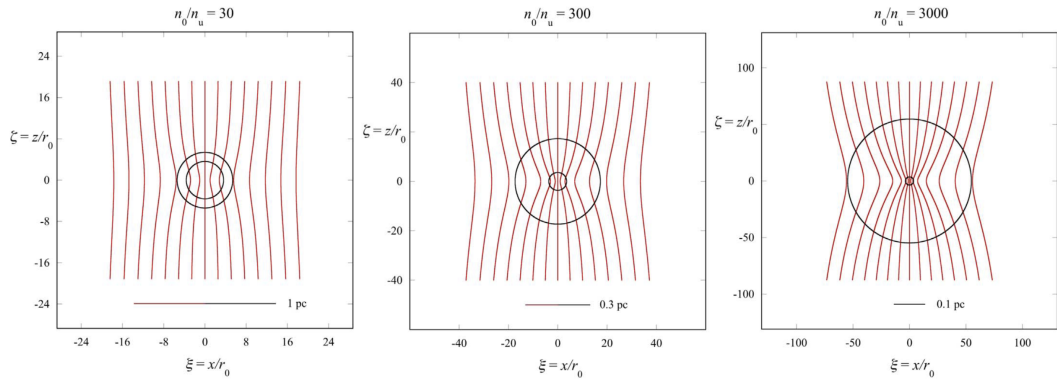
Figure 5.7: *Magnetic flux field lines associated with a $p = 2$ Plummer sphere density distribution, a distribution that is similar to that of the singular isothermal sphere. The central density exceeds the background density by factors 30 (left), 300 (center), and 3000 (right). The inner circle marks a density less than the peak value by a factor of 14.1, as in a critical Bonner-Ebert sphere sphere. The outer circle marks a density with twice the background value. From: Myers et al. ApJ 868, 51 (2018).*

## Magnetic flux freezing and the magnetic critical mass

In a highly conducting medium such as the ISM we can assume *magnetic flux freezing* – that is, that the magnetic field and the gas move together as a single magnetized fluid, or plasma. In other words, the flux lines are 'glued' to the gas. There is evidence that this too holds in the early phases of the collapse of molecular clouds. The 'glueing' of the gas to the magnetic field is done by charged particles in the gas, which gyrate around the field lines at high speeds and couple to the slowly-moving neutral gas through collisions (i.e., the essence of a conducting medium). The ionized particles can move freely along the field lines at the Alfvén speed. In theory this would tend to produce a collapse which proceeds preferentially parallel to the magnetic field, with there being more resistance to collapse perpendicular to the field. Hence collapsing clouds should roughly resemble flattened oblate spheroids, with their long axes perpendicular to the magnetic field. If these oblate spheroids were then to begin to contract perpendicular to their field directions because of increased gravitational attraction as a result of continued mass loading, dragging the field with them (which is the flux freezing), then this would lead to a pinching, or narrowing of the field lines at the centre. Further out along the magnetic field, where material has not been compressed strongly, no such pinching would be seen. This leads to a shape of the magnetic field lines known as the 'hourglass' shape (see figure 5.7).

Hence, if we take a unit volume of that fluid and compress it *orthogonal to the field*, the magnetic field strength increases – i.e. $B \propto 1/V \propto \rho$. Note that this assumes a compressible medium. As the field is not compressed parallel to the magnetic field lines, it is the total flux

threading a given area that is a conserved quantity. Hence

$$B = B_\circ \left( \frac{R_\circ}{R} \right)^2, \tag{5.30}$$

where $R_\circ$ is the initial radius of the cloud and $B_\circ$ the initial magnetic field strength. Therefore, using equation (5.26), within $R$, the magnetic energy is given by

$$E_{\mathrm{mag}} = \frac{1}{8\pi} B^2 \, V = \frac{1}{8\pi} B_\circ^2 \left( \frac{R_\circ}{R} \right)^4 \frac{4}{3} \pi \, R^3 = \frac{B_\circ^2 \, R_\circ^4}{6R}. \tag{5.31}$$

As discussed in section 3.1.5, the magnetic field does not only cause a pressure, it also exerts a tension. Assuming this second term to be of the same magnitude, we may write for the magnetic energy

$$E_{\mathrm{mag}} = \frac{B_\circ^2 \, R_\circ^4}{3R}. \tag{5.32}$$

Using equation (3.55), we require that for equilibrium

$$\frac{B_\circ^2 \, R_\circ^4}{3R} = \frac{3GM^2}{5R}. \tag{5.33}$$

Therefore

$$M_{\mathrm{crit}} = \left( \frac{5}{9G} \right)^{1/2} B_\circ \, R_\circ^2 = \left( \frac{5}{9G} \right)^{1/2} B \, R^2 \tag{5.34}$$

$$\approx 10^3 \left( \frac{B}{30 \, \mu\mathrm{G}} \right) \left( \frac{R}{2 \, \mathrm{pc}} \right)^2 M_\odot, \tag{5.35}$$

where we have employed Eq. (5.30). The mass $M_{\mathrm{crit}}$ is known as the *magnetic critical mass*. The equation shows that it is proportional to $B \, R^2$, which is often referred to as the *magnetic flux*. So indeed, when a magnetic field is present in the cloud, and when there are sufficient charges in the cloud to keep the magnetic field frozen in the gas, magnetic pressure adds to the stability of the cloud.

Notice that the new critical mass, other than the Jeans mass, is independent of the density. So if the magnetic flux in a cloud is sufficient to keep a cloud from collapsing, than no amount of external compression will make the cloud collapse. A magnetic field will thus increase the maximum non-dimensional mass $m$ (see figure 5.4) that indicates the transition from stable to unstable solutions. To initiate collapse it is needed to accrete additional material onto the core to bring the core mass above the critical mass. Once that happens, magnetic pressure is not able to counterbalance the gravitational collapse anymore. So, how does this happen?

### Ambipolar diffusion

When the fraction of ionized particles is very low, the assumption of magnetic flux freezing can break down. The charged species (ions and electrons) in a magnetic medium gyrate around
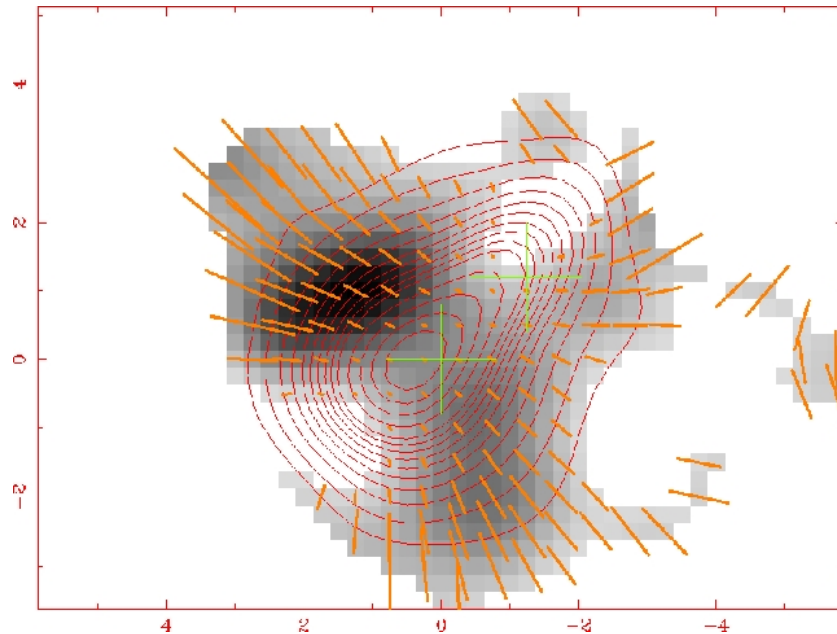
Figure 5.8: *Map of NCG 1333 (or IRAS 4A) at a distance of 300 pc, a typical low-mass stellar system in the earliest stages of evolution, still deeply embedded in an infalling dense molecular and dusty envelope. At this distance 1 arcsec corresponds to 300 AU. Contours show the continuum emission at 877 μm, bars indicate the direction and degree of polarization (tracing the magnetic field vectors), and the color map shows the polarized intensity. Remember that even ionized material can flow along the field lines relatively freely. This would tend to produce a collapse which proceeds preferentially parallel to the magnetic field, with there being more resistance to collapse perpendicular to the field. Hence collapsing clouds should roughly resemble flattened oblate spheroids, with their long axes perpendicular to their field directions; note this by looking at the continuum emission contours. At some point the collapse also proceeds perpendicular to the field directions, dragging the field along. This leads to a pinching, or narrowing, of the field lines at the center. Further out along the magnetic field, no such pinching would be seen. This leads to an 'hourglass-shaped' shape of the magnetic field lines; note this by looking at the polarization bars. This is exactly as predicted from magnetic flux freezing into the gas. From: Goncalves et al. ApJ 490, L39 (2008).*

the field lines with the cyclotron frequency while neutral particles remain unaffected. This sets up relative drift velocities between charged and neutral species leading to frequent collisions that essentially generate the magnetic pressure (and cancel the drift velocities). If the degree of ionization of the medium is however low, neutrals can drift efficiently across the field lines, without feeling a strong pressure from the ions, in a direction imposed by gas pressure gradients or gravitational acceleration. Such differential motion between ions and neutral molecules and atoms is called *ambipolar diffusion*.

For this process to work, and mass to accumulate in the potential well such that the critical mass is exceeded and the cloud collapses, the time $\tau_{\mathrm{AD}}$ for neutrals to move a distance $R$ by ambipolar diffusion should be comparable or shorter than the free fall time $\tau_{\mathrm{ff}}$ (see equa-

tion 3.78).

We follow the reasoning by Ward-Thompson and Whitworth[1] to assess whether this is the case. Consider a configuration where the magnetic field is supporting a gas in a gravitational field. The ions are tied to the field lines and gyrate around them. The neutrals, however, are free to accelerate in the direction of the gravitational force. The velocity reached by a neutral particle (before it collides with an ion and has its velocity randomised) is given by $v \sim g\,t$, where $v$ is the neutral's velocity, $g$ the gravitational acceleration and $t$ is the time between collisions. Under the assumption that the thermal velocities greatly exceed the drift velocity, the collision time can be estimated by averaging the cross-section (which is in general a function of collision energy) and the thermal velocity distribution. Then $t = 1/(n_{ions} < \sigma\,v_{th} >)$, where $n_{ions}$ is the number density of ions, and $< \sigma\,v_{th} >$ is the average over a Maxwellian velocity distribution of the product of the collision cross-section and the thermal velocity.

For a cloud of radius $R$ and gas density $n$ (mainly neutral H and $H_2$), the time for the neutrals to drift past the ions is then the ambipolar diffusion time-scale

$$\tau_{AD} \sim \frac{R}{v} = \frac{R}{g\,t} = \frac{R\,n_{ions} < \sigma\,v_{th} >}{(GM/R^2)} = \frac{3 < \sigma\,v_{th} >}{4\pi G\mu\,m_H}\,q, \tag{5.36}$$

where $\mu$ is the mean molecular weight and $q = n_{ions}/n$ is the ionization fraction. Note that the diffusion time-scale depends only on the fractional ionization and the temperature of the gas (through the average Maxwellian velocity). Inserting values appropriate for $T = 10\,\mathrm{K}$ yields $\tau_{AD} = 7.3 \times 10^{13}\,q$ years. Thus ambipolar diffusion is a very slow process, except in regions of very low ionization.

Dark interstellar clouds are effectively shielded from the background UV interstellar radiation field because of the high (internal) dust extinction. Ions inside the cloud are mainly produced by comic ray (CR) ionization. The number of cosmic ray ionizations is roughly $I_{CR} \sim 10^{-17}\,n$ cm$^{-3}$s$^{-1}$. The number of recombinations depends on the density of the electrons $n_e$ and is about $R_R = 8.2 \times 10^{-13}\,n_{ions}\,n_e$ cm$^{-3}$ s$^{-1}$. Balancing the two rates and assuming $n_{ions} \simeq n_e$ gives

$$q = 3.5 \times 10^{-8} \left(\frac{10^4\,\mathrm{cm}^{-3}}{n}\right)^{1/2}, \tag{5.37}$$

and hence the ambipolar diffusion time-scale (for a gas at $T = 10\,\mathrm{K}$) is given by

$$\tau_{AD} = 2.5 \times 10^6 \left(\frac{10^4\,\mathrm{cm}^{-3}}{n}\right)^{1/2}. \tag{5.38}$$

Comparing this result with that of the free-fall time Eq. (3.79) shows that $\tau_{AD}$ typically is about an order of magnitude larger than $\tau_{ff}$ and that it thus, generally, does not seem to be an important effect.

---

[1] Ward-Thompson and Whitward, in An Introduction to Star Formation. Be ware, on the basis of this reasoning we reach the opposite conclusion!

In cases of exceptionally low ionization, ambipolar diffusion may however occur. In those cases, the inward drift of the neutral particles (because of gravitational pull) increases the central density, thus decreases the time-scale $\tau_{\mathrm{AD}}$. This has an important effect on the critical mass, as defined by equation (5.35). As the neutrals drift past the ions and the field lines, the radius of the cloud (initially $R_\circ$) decreases but the magnetic field (initially $B_\circ$) does not change. The magnetic flux, here $B_\circ R_\circ^2$, thus decreases, and hence the magnetic critical mass also decreases. In fact, the magnetic flux (and hence the magnetic critical mass) will tend to zero in a time of order $\tau_{\mathrm{AD}}$. So, unavoidable, then, the actual cloud mass must at some point exceed the magnetic critical mass, and the cloud will then start to collapse.

Thus the entire effect of magnetic fields on the equilibrium of clouds can be summarized as follows:

- ○ Density perturbations that exceed the magnetic critical mass will undergo collapse immediately.

- ○ For sub-critical masses which are sufficiently dense to have an exceptionally low degree of ionization, the magnetic field will hold up collapse for a time given by $\tau_{\mathrm{AD}}$, after which collapse occurs.

- ○ For sub-critical clouds of sufficiently low density and consequently high ionization (because the cloud can not efficiently block the UV background field), the magnetic field can support the cloud almost indefinitely.

### The angular momentum problem and magnetic braking of rotating clouds

In section 5.2 we discussed rotational support and concluded that for typical initial cloud radii rotational energies are of the order of 1 percent of the gravitational energy (see equation 5.25), which is insufficient to stabilize the clouds. However, angular momentum, i.e. $I\,\omega$ where $I$ is the moment of inertia and $\omega$ is the angular velocity, is a conserved quantity. This implies that

$$M\,R^2\,\omega = M\,R_\circ^2\,\omega_\circ. \tag{5.39}$$

Thus as the cloud collapses, $R$ decreases, $\omega$ increases, and the rotational kinetic energy increases until it may balance the gravitational energy. In that case a rotationally supported cloud results. Using (5.39) to rewrite (5.25) yields

$$\frac{E_{\mathrm{rot}}}{E_{\mathrm{grav}}} = \left(\frac{R_\circ}{R}\right)\left.\frac{E_{\mathrm{rot}}}{E_{\mathrm{grav}}}\right|_\circ \tag{5.40}$$

For equilibrium this ratio should be ∼unity. With typical molecular cloud cores rotating at $E_{\mathrm{rot}}/E_{\mathrm{grav}} \sim 0.01$, the cloud radius need only decrease by a factor of about 100 before rotation supports the cloud. A typical cloud might start with a radius of 0.1 pc ($3 \times 10^{17}$ cm). After it has contracted by a factor of 100 it would still have a radius of $3 \times 10^{15}$ cm or 200 AU. Compared to the solar radius of $\sim 7 \times 10^{10}$ cm, the core radius would be fully supported

by rotation some five orders of magnitude in size away from forming an object of stellar size. This angular momentum problem is a serious impediment to star formation. Magnetic braking of rotating clouds may solve this problem.

Magnetic fields can be an important mechanism for carrying away angular momentum from cloud cores. Intuitively, one may expect this since the magnetic fields emerging from a rotating cloud core are tied like rubber bands to the larger Galactic magnetic field, and so would be expected to produce a torque as they are wound up[2]. As the cloud rotates it starts to wind-up the field lines. The ions that gyrate around the magnetic field lines can travel quite easily along these lines (with a speed that is called the Alfvén speed) and propagate this rotation, i.e. spin-up the lower density gas outside the cloud itself. Since angular momentum must be conserved this results in a slowing down of the cloud's rotation.

---

[2]Note that this assumes that the rotation axis is parallel to the magnetic field direction. If they are not parallel to one another, a more complex geometry ensues.

## Exercise 5.1

a) Use equations (5.1), (5.2), and (5.5) to show that the solution of the density structure for infinite central density is given by equation (5.6).

b) Show that the mass up to radius $r$ for the singular isothermal sphere is given by equation (5.7).

## Exercise 5.2

A solution which satisfies the Lane-Emden equation (5.15), *but not the boundary conditions at $\xi = 0$*, is

$$\Psi(\xi) = \ln\left(\frac{\xi^2}{2}\right),\tag{5.41}$$

and is called the isothermal singular sphere. Verify this solution and show that the density structure of this solution is given by equation (5.6).

## Exercise 5.3

Verify that $m_\infty = (2/\pi)^{1/2}$ represents the non-dimensional mass of the singular isothermal sphere.

## Exercise 5.4

Derive an expression for the Bonnor-Ebert mass similar to the last right-hand side of expression (3.61) for the Jeans mass.

## Exercise 5.5

A typical cloud contains much more angular momentum than can eventually be stored in a central stellar object. The final sentence of the chapter states that magnetic breaking may remove angular momentum from collapsing rotating gas clouds. This must occur as it is impossible for a central stellar object to store all of this angular momentum in its rotation – the spin velocity that is required to do so would vastly exceed the break-up speed. Think of other methods that may help to avert loading a central object with too much angular momentum.

# 6

# Building a typical star

For a core to condense into a star requires a vast change of density and temperature. Relative to the conditions of the initial cloud, we need to achieve an increase in density by 20 orders of magnitude and an increase in temperature by 6 orders of magnitude to form a star. In this chapter, we follow the evolution from a collapsing core in a molecular cloud to a newly formed star as it approaches the main sequence.

## 6.1  Inside-out collapse

In section 3.4 we discussed pressure-free collapse (leading up to an expression for the free-fall time), pointing out that in that case the collapse was homologous, i.e. the density increased at the same rate everywhere in the cloud. If we assume that the central parts of the core have higher densities than the outer parts, as suggested by solutions of hydrostatic equilibrium discussed in the previous chapter, the collapse will be highly non-homologous. In that case, the free-fall time $\tau_{\text{ff}} \propto \rho^{-1/2}$ increases with distance. This means that the central parts collapse faster than the outer parts and the outer shells subsequently loose their pressure support, while the mass and hence the gravitational attraction inside such a shell does not change. The information of the 'lost pressure support' travels outwards as a wave with the sound speed $a$. At this rate, the region of infall slowly spreads into static, more rarefied gas.

In 1977, Frank Shu (ApJ 214, 488) developed the so-called inside-out collapse scenario and a self-similar dimensionless solution to the problem. Later numerical simulations identified four distinct phases:

(1) **Free-fall or isothermal collapse phase.** The collapse is approximately isothermal and the timescale is the free-fall timescale. Efficient cooling is provided by molecular hydrogen and thermal radiation by dust grains. The free-fall or isothermal collapse phase ends with the formation of an opaque, hydrostatic object at the centre, surrounded by a gaseous envelope.

The next two phases are *adiabatic collapse phases*, because the inner parts of the core become optically thick and thus cannot cool as efficiently anymore.

(2) **First core phase.** Because the material in the center becomes optically thick it is more difficult for energy to escape, hence the material heats up. As heating progresses, the force produced by the pressure gradient decelerates material falling inward and the collapse slows down in a maturing shock front. A small amount of mass in the center (of order $1 - 5 \times 10^{-2} \, M_\odot$) approaches hydrostatic equilibrium; this region is known as the *first stable core*. The radius of the first core is large by stellar standards, roughly a few AU (let us say roughly 5 AU). With addition of mass and shrinking of the radius, $T$ first surpasses 1000 K, where dust grains evaporate removing them as an opacity source, and later 2000 K, where collisional dissociation of $H_2$ begins. The dissociation of molecular hydrogen requires energy (4.52 eV per bond breaking) and this severely damps the rise of the interior temperature. With the temperature only slowly rising any further compression can no longer be effectively opposed by a rise in the internal pressure, and the entire configuration becomes unstable and collapses. This event marks the end of the first core.

(3) **Second core or opacity phase.** The collapse of the partially dissociated gas takes the central region to much higher density and temperature. As a result of the rapidly increasing pressure at the center, the material in this second collapse regains equilibrium after the dissociation of $H_2$ has been completed at a central temperature of $\sim$20 000 K. The density by this time has increased to about $10^{-2} \, \mathrm{gr \, cm^{-3}}$, so the physical conditions are approaching those in stars; however the initial mass of this *second core* is at first small, only some $10^{-3} \, M_\odot$. The *second hydrostatic core* also develops a shock front at the edge, which has a radius of only a few $R_\odot$, and it gradually builds up mass by accretion through the shock front. As the core temperature increases, the hydrogen begins to ionize.

The final phase is characterized by the entire core being optically thick and overall accretion.

(4) **Accretion phase** A radius of several solar radii and a mass of $\sim$0.1 $M_\odot$ brings the mean temperature, using equation (6.16), above $10^5$ K. Such a value, coupled with a mass density of order $10^{-2} \, \mathrm{gr \, cm^{-3}}$, places the object within the stellar regime. By this point, the central object is said to have entered the *main accretion phase*. At first, most of the gravitational energy goes into dissociating in-falling $H_2$ and ionizing hydrogen and helium, i.e. in increasing the internal energy, and only some is radiated into space. Later, this situation is reversed. The central object becomes a *proto-star*, defined as an object whose luminosity stems mainly from external accretion. In this phase, a *proto-planetary disk* forms. Actually, most of the material is accreted onto the proto-star through this disk. It is also in this phase that the proto-star becomes first visible in the infrared. Low mass ($M \stackrel{<}{\sim} 2 M_\odot$) stars finish their accretion – and thus become

visible – prior to reaching the main sequence. They continue to contract until their central temperature becomes high enough for deuterium and, somewhat later, hydrogen burning to start. Contraction is (again) responsible for their luminosity, though these *pre-main sequence stars* may still have a proto-planetary disk and a low accretion rate. More massive stars often start hydrogen burning before the accretion phase ends.

In the next sections we discuss aspects of this 'road to star formation' in somewhat more detail. Note that these discussions will be simplified as the transition from a more-or-less spherical accretion to a disk-mediated accretion is still out of reach for even for the most sophisticated modeling, but see Section 4.5.

## 6.2 Free-fall or isothermal collapse phase

Once the cloud exceeds the critical Bonnor-Ebert mass (equation 5.21), no hydrostatic state of equilibrium is accessible to the system anymore, and the cloud must evolve dynamically. It is then to be described by the equations of fluid dynamics. In spherical geometry the relevant equations are (see equations 3.3, 3.10, and 3.23)

$$\frac{\partial M(r)}{\partial t} + v\frac{\partial M(r)}{\partial r} = 0, \qquad \text{or} \qquad \frac{\partial M(r)}{\partial r} = 4\pi r^2 \rho \qquad (6.1)$$

and

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial r} = -\frac{a^2}{\rho}\frac{\partial \rho}{\partial r} - \frac{GM(r)}{r^2}, \qquad (6.2)$$

where $M(r)$ is the mass inside the radial distance $r$ and $v$ de infall velocity. One should note that Eq. (3.15) should actually have been defined with a minus sign, as for a negative velocity (i.e. material is flowing in) the mass $M(r)$ is increasing.

Shu (1977) presents a discussion of solutions for this set of equations, applying different boundary conditions. He notices that if one concentrates on the part of the flow that is well removed from the outer boundary and assumes that the collapse occurs over many orders of magnitude in density, so that the central object has negligible dimensions in comparison with the region of accretion flow, the only quantities of the problem are $G$, $a$, $r$, and the time $t$. We can understand the inside-out collapse as an outward traveling 'expansion wave', i.e. the collapse front moves out in time and at time $t$ is at position $at$ (see Figure 6.1). If we define a dimension-less radial variable

$$x = \frac{r}{at}, \qquad (6.3)$$

such that in the inside-out collapse $x = 1$ always corresponds to the position of the expanding wavefront, one may look for (similarity) solutions of the form

$$\rho(r,t) = \frac{\alpha(x)}{4\pi G t^2}, \qquad M(r,t) = \frac{a^3 t}{G}\,m(x), \qquad v(r,t) = a\,u(x). \qquad (6.4)$$
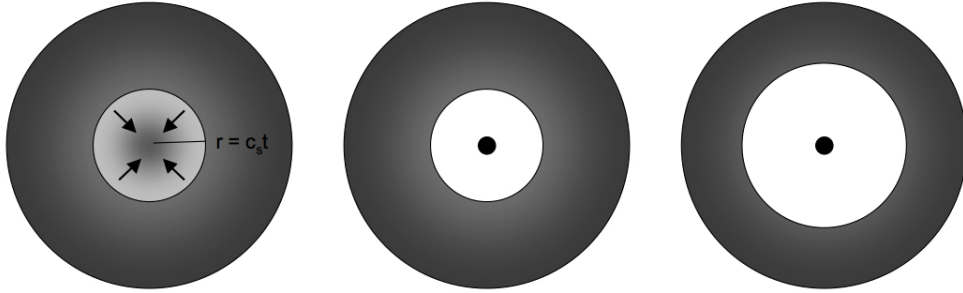
Figure 6.1: *Sketch of the inside-out collapse scenario of Shu. The expansion wave travels outward at a speed $a$ (in the figure named $c_s$) and material inward of the wavefront collapses on the central object.*

The dimension-less functions $\alpha(x)$, $m(x)$, and $u(x)$ need to be determined (for given boundary conditions).

### Static solution

For the static case, the solution is given by $u = 0$, $\alpha = 2/x^2$, and $m = 2x$. In that case one recovers the singular isothermal sphere, with $\rho(r)$ given by equation (5.6) and the mass $M(r)$ by equation (5.7).

### Collapse solution featuring a critical point

Another apparent solution of the equation of fluid dynamics is given by $x - u = 1$ and $\alpha = 2/x$. The point $x = 1 + u$ is a *singular point* of the ordinary differential equations (6.1) and (6.2). We do not elaborate here on this type of solution, except for stating that for the flow to pass smoothly through such a singular point, $\alpha$ must equal $2/x$ where $x = 1 + u$. For such solutions, nothing singular happens at the singular point, and the 'singular point' is more appropriately referred to as a *critical point*. The situation appears similar to Bondi's (1952) discussion of the critical accretion rate in a steady spherical-accretion flow, but is actually not (because for the time-dependent problem critical points can not generally represent the gravitational collapse of isothermal spheres).

In this type of solution the material accumulates at $x = 0$, which is of course unphysical. The real central object has a finite size, hence, the solution should be terminated at some small value of $x$ by introducing a shock. We shall however not discuss this point further.

### Expansion-wave or Inside-out collapse solution

General inside-out collapse solutions, with an expansion wave traveling outward obeying Eq. (6.3), require a numerical treatment. Shu provides a table with values $\alpha(x)$, $m(x)$, and
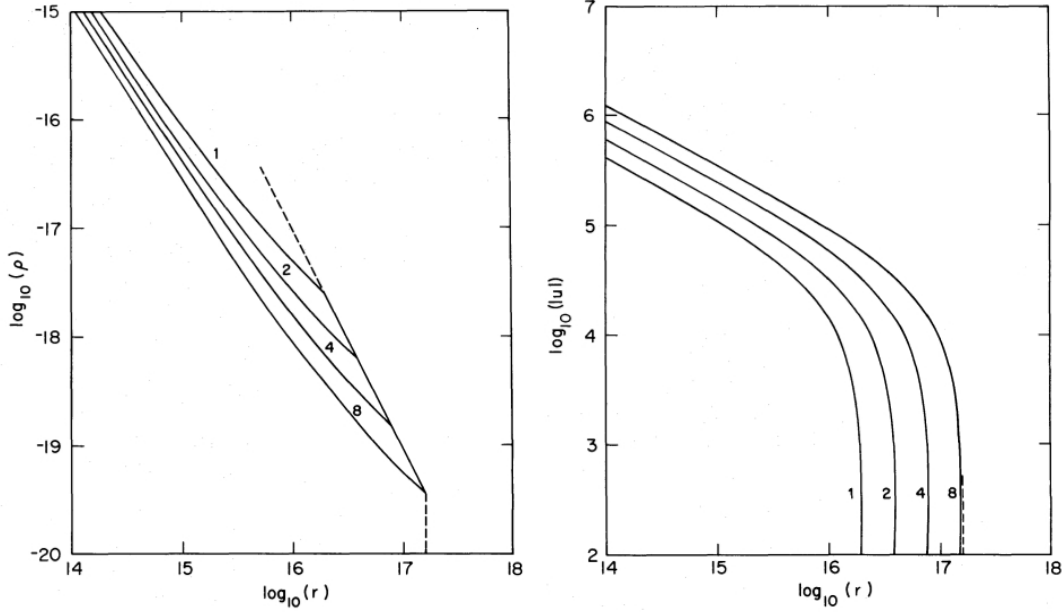
Figure 6.2: *Left: Density as a function of radial coordinate for the inside-out collapse of a 0.96 M$_\odot$ core with an initial velocity of 0.2 km/s. The initial radius of the outer boundary is indicated by the dashed line. The different profiles correspond to $t = 1, 2, 4, 8 \times 10^{12}$ s (or $3.2, 6.3, 12.7, 25.4 \times 10^4$ yr). Right: The corresponding velocity profiles. The radius is in cm and the density in gr cm$^{-3}$.*

$u(x)$. The solution for $0 \le x \le 1$ can approximately, but very accurately, be described by

$$
\begin{aligned}
\alpha(x) &= \left(g(x)^{7/2} + h(x)^{7/2}\right)^{2/7} \\
m(x) &= 1.025\,x^2 + 0.975 + 0.075\,x\,(1-x) \\
u(x) &= \left(h(x)^{5/9} - 2^{5/9}\right)^{9/10},
\end{aligned}
\tag{6.5}
$$

where $g(x) = 1/\left(1.43\,x^{3/2}\right)$ and $h(x) = 2/x$. For any time $t$ this can be converted into a real solution. The inner part of the density profile is determined by the free-fall collapse. For this regime (i.e. $x \ll 1$ or $r \ll at$), the density and velocity can be described as

$$
\rho(r,t) = \frac{a^{3/2}}{17.96\,G}\frac{1}{\sqrt{t}}\frac{1}{r^{3/2}} \qquad \text{and} \qquad v(r,t) = \sqrt{\frac{2GM_\star(t)}{r}}, \tag{6.6}
$$

where $M_\star(t)$ is the mass of the central object at time $t$. The free-fall velocity $v(r,t)$, we might call it $v_{\mathrm{ff}}$ for convenience, at given time $t$ simply describes that potential energy is converted into kinetic velocity when the gas falls toward mass $M_\star$, i.e. it follows immediately from equating $E_{\mathrm{grav}} = GM_\star m/r$ to $E_{\mathrm{kin}} = mv^2/2$. The mass accretion rate of the sink-cell mass at the center, i.e. at $x = 0$, is constant and given by

$$
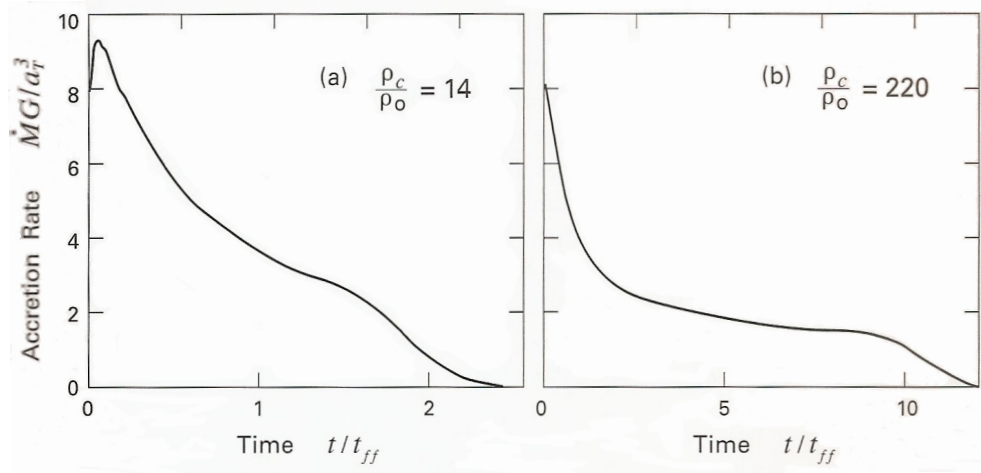\dot{M} = 0.975\,\frac{a^3}{G} \tag{6.7}
$$

Figure 6.3: *Evolution of the mass accretion rate in two collapsing, isothermal spheres. Time is measured after the start of proto-star formation, and is in units of the free-fall time. Each cloud has the initial density contrast indicated.*

$$\simeq\; 1.6 \times 10^{-6} \left( \frac{T}{10\,\mathrm{K}} \right)^{3/2} M_\odot \mathrm{yr}^{-1}. \tag{6.8}$$

This implies that the mass grows linear in time and that therefore a star such as the sun accumulated its mass in only $\sim$ half a million years (see Eq. 6.20). Note that a constant accretion rate is indeed what one may intuitively expect. For instance, a dimensional analysis of the rate at which the central object accretes mass in the free-fall phase can be obtained using equations (3.3), (3.61), and (3.78), and assuming $\dot{M} \propto M/\tau_{\mathrm{ff}}$. This yields a constant mass-accretion rate $\dot{M} \propto a^3/G$.

The solution for a $M_\star = 0.96\, M_\odot$ object is graphically shown in Figure 6.2. Close to the star the material is in free fall and $\rho \propto r^{-3/2}$. As we move outwards in the solution, we encounter a transition region, where matter starts to fall towards the center. The kink in the profile indicates the position of the expansion wavefront (again, always at $x = 1$). At this radius the infall velocity $u = 0$. Outside this point we see an isothermal sphere solution with $\rho \propto r^{-2}$.

## Hydrodynamical solutions

Numerical simulations of the hydrodynamical evolution of the mass accretion rate onto the sink-cell mass for two collapsing, isothermal spheres are shown in Figure 6.3 for two starting values of the central density contrast. These solutions differ from the Shu inside-out collapse solution in that they do not mandate $x = 1$ and require (or apply) a perturbation of the outer-boundary condition to initiate collapse. The left panel is for the density contrast of almost the Bonnor-Ebert sphere, which without a perturbation would be stable. Here, however, a

perturbation is applied. The right panel is for a larger central density contrast, therefore this cloud is intrinsically unstable and we must view this result with caution. Still, we focus on the latter case. The initial behavior of $\dot{M}$ is sensitive to the precise starting configuration (including the manner of the perturbation), but drops rapidly at first, leveling of to a value of $m(t)$ (which is what is plotted on the vertical axis) of about 2 over a longer period before the final truncation. This final truncation represents the collapse of the cloud boundary into the sink cell. Since the surrounding medium is assumed to have zero density, $\dot{M}$ vanishes.

## 6.3 The adiabatic collapse phases: the first and second core phase

At the beginning of the isothermal collapse the opacity from dust grains, $\kappa$, is below 1 cm$^2$ gr$^{-1}$; the density is about $10^{-19}$ gr cm$^{-3}$, and the size of the cloud is about $10^{17}$ cm. This implies that the optical depth from center to edge $\tau \approx \kappa \rho R \ll 1$. The cloud can easily remove energy by radiation to keep the temperature from rising. The free-fall phase will end once the central lump becomes opaque to its own (cooling) radiation. Further compression then causes its internal temperature to rise steadily and the growing gas pressure acts to stabilize the core (by establishing hydrostatic equilibrium). Let us first briefly discuss the nature and properties of the opacity.

### Opacities

An overview of the behavior of the opacity as a function of temperature and density is given in Figure 6.4. It is dominated by dust grains at temperatures up to about 1500-2000 K, by molecules and atoms up to about 10 000 K, and by ions above that temperature range. Between temperatures of about 400 and 1500 K the grains are composed of mainly silicates, iron, and their compounds. Between 400 and 175 K, volatile organics are also present. Below 175 K, ice grains are present along with the other species and are actually dominating. In this low temperature range the Rosseland mean opacity $\kappa_R$ increases as $T^2$. Between 175 K and about 1700 K the opacity is dominated by silicates and iron grains. The sudden dip at about 1700 K is caused by the evaporation of the silicate and iron grains; above that temperature and up to about 3500 K molecular opacity dominates with important contributors being $H_2O$, $CH_4$, $NH_3$, and CO. In between $\sim 3\,500$ K and 9 000 K, the $H^-$ anion is the main opacity source (for instance, in the solar photosphere), while at higher temperatures first hydrogen bound-free and free-free processes take over, supplemented by electrons scattering and He b-f and f-f processes at about 15 000 K and up.

Calculating the Rosseland mean opacity for a gas mixture is very challenging as it requires knowledge of the individual opacities of tens to hundreds of millions of atomic and, particularly, molecular transitions from neutral and ionized stages. Much of the pioniering work in this field has been done in relation to nuclear bomb experiments after the second world war. Later on, astrophysics benefitted from this data. Notable large efforts have been the Opacity
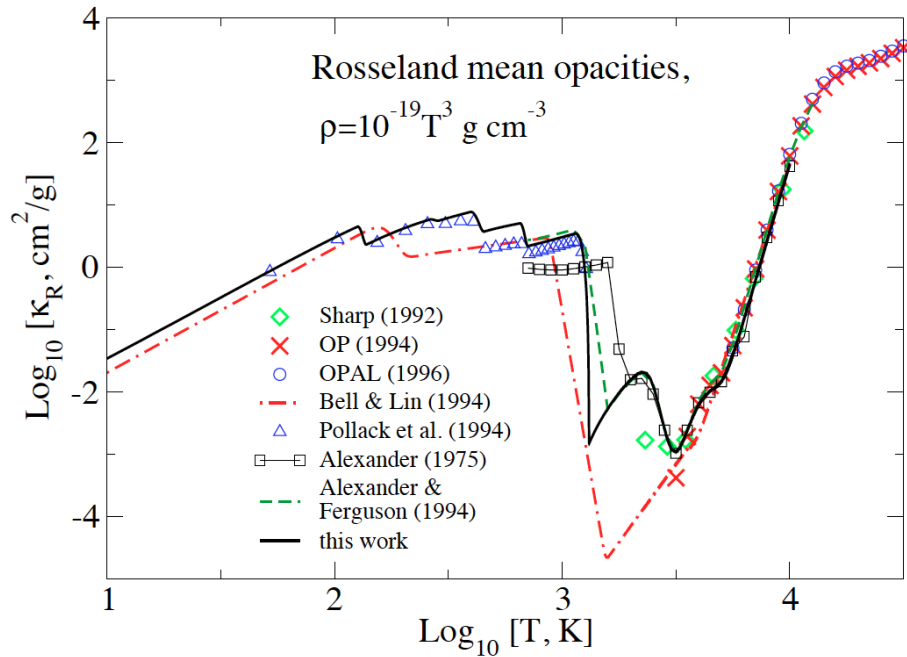
Figure 6.4: *The Rosseland mean opacities in $cm^2\,gr^{-1}$ calculated for temperatures between 1 K and $10^5$ K and gas densities $\rho = 10^{-19}\,T^3\,gr\,cm^{-3}$. The adopted C/O abundance is 0.43. Dust grains dominate the opacity up to a temperature of about 1500-2000 K; atoms dominate beyond 4000 K. The dust grains consist of a mix of silicates, organics, amorphous ice, FeS, and iron. The ice grains evaporate at about 160 K. The results of Semenov et al. are compared to opacity tables provided by the Opacity Project (OP; 1994, crosses), the Opacity Project at Livermore (OPAL; 1996, circles) and other, smaller, projects. From: Semenov et al. 2003, A&A 410, 611.*

Project (OP) and the Opacity Project at Livermore (OPAL). Results of these projects are compared to the work of Semenov et al. (2003) in the overlapping part of the parameter domain in Fig. 6.4.

## Instabilities due to changes in the state of the gas

In the early adiabatic phase changes in the state of the gas may lead to instabilities. There are several ways in which these instabilities may arise. Here we discuss that an unstable system may result when $\gamma$ becomes too low (less than 4/3 to be specific). Note that a high value of $\gamma$ may also have effects. For instance, for $\gamma > 5/3$ the systems would become unstable to convection and energy transport could happen that way as well (as opposed to radiative transport only). The change of $\gamma$ (within the allowed bounds) may itself lead to an instability as it, essentially, describes how strongly an adiabatic gas resists being compressed. That is, it specifies how rapidly the pressure rises in respons to an increase in density. The larger the value of $\gamma$, the harder it is to compress a gas. It thus has a profound influence on the pressure (structure) of the developing core.

Instabilities may also arise because of rapid changes in the opacity of the material in connection with increasing temperature and density (and are generally referred to as *secular instabilities*). If $\kappa_R$ increases upon adiabatic compression, we call these instabilities *vibrational instabilities*.

Instabilities may also occur when the energy that needs to be invested to change the state of the gas is large compared to the thermal energy of the particles. This is the case for the collisional dissociation of $H_2$, and we will discuss this shortly.

As eluded to above, an adiabatic situation requires $\gamma > 4/3$ for the system to be bound or stable. We can show this in the following way: for an adiabatic process the pressure $p$ and the volume $V$ of a gas are related according to Equation 3.4, which implies $pV^\gamma$ is a constant. This means that for a small adiabatic change in volume and pressure, we can write

$$\gamma \frac{dV}{V} + \frac{dp}{p} = 0 \tag{6.9}$$
$$V\,dp = -\gamma\,p\,dV,$$

and hence

$$d(pV) = p\,dV + V\,dp \tag{6.10}$$
$$= -(\gamma - 1)\,p\,dV.$$

During an adiabatic collapse, there is no heat transfer to or from the ambient medium and so the change in internal energy $E_{in}$ of the system equals the work done on the system. We can thus write

$$dE_{in} = -p\,dV = \frac{1}{\gamma - 1}d(pV), \tag{6.11}$$

where the minus signs enters because a compression implies a negative change of the volume. Hence, for a constant adiabatic index $\gamma$, the relation

$$E_{in} = \frac{1}{\gamma - 1}\,pV \tag{6.12}$$

holds. For a system in virial equilibrium, the average pressure $< p >$ can be written as one third of the density of stored gravitational energy (see equations 3.42 and 3.44), i.e.

$$< p >= -\frac{1}{3}\frac{E_{grav}}{V}. \tag{6.13}$$

For a self-gravitating gas in hydrostatic equilibrium, we can thus write

$$-\frac{1}{3}\frac{E_{grav}}{V} = (\gamma - 1)\frac{E_{in}}{V}$$
$$3(\gamma - 1)E_{in} + E_{grav} = 0 \tag{6.14}$$

The total energy is

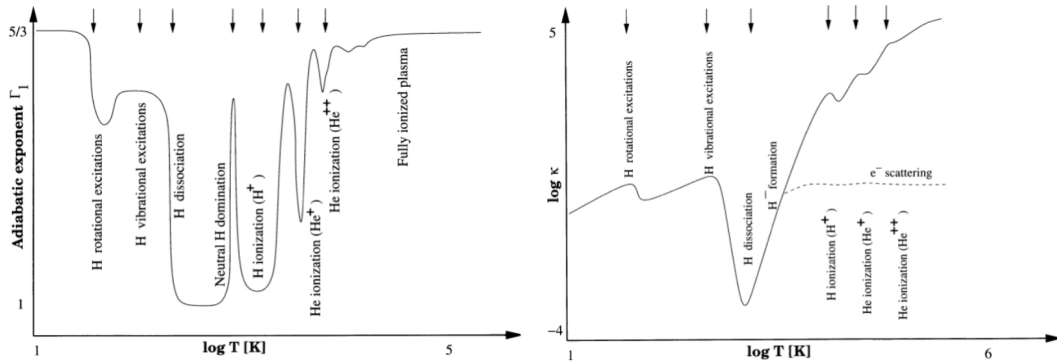$$E_{tot} = E_{in} + E_{grav} = (4 - 3\gamma)E_{in}, \tag{6.15}$$

Figure 6.5: *Left: The adiabatic index* $\gamma = \Gamma_1 = (\partial \ln p / \partial \ln \rho)_{\rm ad}$, *also called the first adiabatic index, as a function of temperature. The arrows mark the onset of specific physical processes that change the index most dramatically causing instability in the nearly formed stellar core. Right: The Rosseland mean opacity* $\kappa_R$ *as a function of temperature taken from the same calculation as the adiabatic index. It shows the behaviour of opacity parallel to the changes in the adiabatic index and thus how it can counter drifts into instability regions. Comment: The leftmost three arrows in both panels should be labelled with* $H_2$ *instead of H. From: Schulz 2005.*

and the system is bound or stable if $E_{\rm tot} < 0$, hence if $\gamma > 4/3$.

Fig. 6.5 shows on the left the adiabatic index as a function of temperature $T$. With increasing temperature, the molecules in the gas get first excited, and then they dissociate ($H_2$ at $\sim$ 2000 K[1]). Above a critical temperature, which is density dependent, the gas becomes more and more ionized. The figure was computed for a gas density of $10^{-14}$ g cm$^{-3}$. With changing density, the depth and heights of these peaks and valleys in $\gamma$ change.

### Early growth and collapse of the first core

At some point in time, the accumulated matter at the center of the cloud will form a core that is optically thick. Let us estimate the mean internal temperature of this core, utilizing the virial theorem in the version of equation (3.46), i.e. neglecting rotational and magnetic support – the idea being that the very center of the collapsing core is least supported rotationally and magnetically. We approximate the gravitational energy as $-3GM^2/5R$, for a core mass $M$ and radius $R$. The internal energy $E_{\rm th} = 3NkT/2$ (see equation 3.47) becomes $3\mathcal{R}\,T\,M/2\mu$, using $M = \mu\,m_{\rm H}N$, and where $T$ and $\mu$ are the volume-averaged temperature and molecular weight, respectively. Applying equation (3.46) and solving for the temperature, we find

$$T \sim \frac{\mu}{5\mathcal{R}} \frac{G\,M}{R} = 510 \left( \frac{M}{5 \times 10^{-2}\,M_\odot} \right) \left( \frac{R}{5\,AU} \right)^{-1} \quad {\rm K}, \qquad (6.16)$$

---

[1]Under conditions of thermodynamic equilibrium, the dissociation constant $K(H_2)$ is related to the partial pressure of the atomic constituent $p(H)$ and $p(H_2)$ through $K(H_2) = p^2(H)/p(H_2)$. The dissociation constant $K$ depends only on temperature.

where we have set $\mu$ equal to 2.4, the value appropriate for molecular gas.

Outside the core, infall velocities become supersonic with respect to the slowly compressing core material, and a shock front forms at its outer edge, at a radius of a few AU (let us say at $\sim$5 AU). So, once $5 \times 10^{-2} \, M_\odot$ has accumulated in the central core, the temperature has risen to $\sim$500 K. This temperature is still very low compared to true stars, but is certainly much higher than in quiescent molecular clouds, as is the average density, which is now of order $10^{-10} \, \mathrm{gr\,cm}^{-3}$.

At temperatures in excess of 1000 K, dust grains evaporate removing them as an opacity source at temperatures of about 1500-2000 K. Before the core has had a chance to accumulate much mass, it compresses to a density of $10^{-8} \, \mathrm{gr\,cm}^{-3}$ and $T \sim 2000$ K, where $H_2$ begins to dissociate.

Let us compare the thermal energy of molecular hydrogen with the dissociation energy. The internal energy of molecular hydrogen is

$$E_{\mathrm{th},H_2} = \frac{f}{2} \, n \, k \, T = \frac{f}{2} \, \frac{\rho}{\mu \, m_\mathrm{H}} \, k \, T = \frac{f}{2} \, \frac{\mathcal{R} \, \rho \, T}{\mu}, \tag{6.17}$$

where $f$ denotes the degrees of freedom. For $T < 80$ K only translational degrees of freedom are excited and $f = 3$. Above that temperature, rotational degrees of freedom begin to become excited, and in the range $300 < T < 1000$ it holds that $f = 5$. Between 80 and 300 K, the degree of excitation must be calculated in detail. Above 1000 K the vibrational levels also become excited, contributing additional degrees of freedom; however the dissociation of the molecules occurs before this excitation is complete.

The dissociation energy per $H_2$ molecule is 4.48 eV, while the thermal energy of a molecule is $5kT/2 \sim 0.42$ eV at 2000 K. The dissociation energy per molecule is so large compared with its thermal energy that most of the gravitational energy released on compression goes into dissociation energy rather than into heating of – and supporting – the gas.

Let us describe the events up to this point and what happens next using the schematic diagram of Figure 6.6. In doing so, we closely follow Bodenheimer (2011). The initial horizontal portion of the diagram is the isothermal phase. The material becomes optically thick and begins to heat near point $A$ in the figure. The heating may be described by Eq. (6.16). After a short period of readjustment, during which the rotational degrees of freedom are excited in the $H_2$ molecules, the collapse approaches an adiabat. In that case, $d \log T / d \log \rho = (\gamma - 1)$ from Eqs. (3.3) and (3.4). The slope in the diagram follows $d \log T / d \log \rho \approx 0.42$; for pure molecular hydrogen with 5 degrees of freedom it would be 0.4. As heating progresses, the force produced by the pressure gradient counters that of gravity, and the collapse slows down. Near point $B$ in the diagram a small amount of mass in the center ($\sim 10^{-2} M_\odot$) approaches hydrostatic equilibrium; this is the *first core*. It starts with a temperature of a few hundred Kelvin and a density of about $10^{-10} \, \mathrm{gr\,cm}^{-3}$. Outside the core, infall velocities become supersonic with respect to the slowly compressing core material, and a shock wave forms at its outer edge, at a radius of about 5 AU.
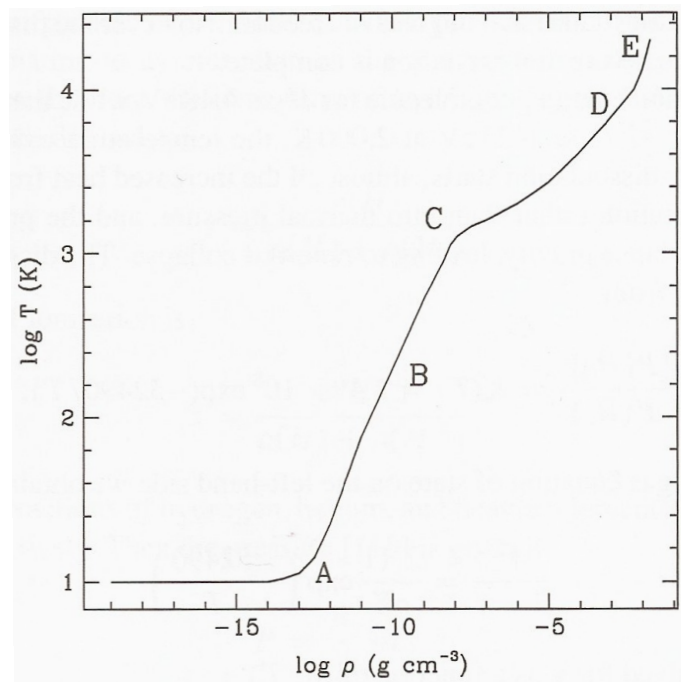
Figure 6.6: *Schematic diagram of the evolution of the center of a proto-star during the isothermal phase (horizontal portion of the curve) and the adiabatic phase. Salient points, as described in the text, are marked by letters. The curve represents the transition of gas from an interstellar cloud to a star. From: Bodenheimer (2011), Principles of Star Formation.*

Before the core has had a chance to accumulate much mass, it compresses to a density of about $10^{-8}$ gr cm$^{-3}$ and $T = 2000$ K, where H$_2$ begins to dissociate (point C). The dissociation energy per molecule is so large compared with its thermal energy (see above) that most of the gravitational energy released on compression goes into dissociation energy rather than into heating of the gas. The increase in temperature slows noticeably, and $\gamma$ falls to about 1.1 (i.e. below 4/3). The collapse, temporarily halted in the first core, begins again in the center of that core, and velocities eventually approach free fall. Collapse continues over several more orders of magnitude in density, until the H$_2$ is mostly dissociated in the very center (about 8000 K; point D). The curve steepens, approaching $d \log T / d \log \rho = 0.67$, as would be expected for a gas of neutral H and He with $\gamma = 5/3$.

As a result of the rapidly increasing pressure at the center, the material in this second collapse regains equilibrium after the dissociation of molecular hydrogen has completed and the central temperature has increased to 20000 K (point E). The density by this time has increased to about $10^{-2}$ gr cm$^{-3}$, so the physical conditions are approaching those in stars; however, the initial mass of the stellar core is only $10^{-3}$ $M_\odot$. Most of the mass of the proto-star at this time is still in the outer isothermal region, at densities of $10^{-18} - 10^{-19}$ gr cm$^{-3}$, 17 orders of magnitude lower than in the center.

The *second hydrostatic core* also develops a shock front at the edge, which has a radius of only a few $R_\odot$, and it gradually builds up by accretion through the shock front. The first shock front disappears relatively soon as it erodes because of the collapse occurring at its center. As the core temperature increases, the hydrogen begins to ionize, at temperatures considerably higher than those in the ionization zones of stellar atmospheres, because of the much higher density here. One might expect that a third collapse would occur because of the hydrogen ionization (which requires energy). However, the increase in degree of ionization with temperature is much slower than in low-density atmospheres, and furthermore the equation of state under these conditions is beginning to be non-ideal, that is, interactions between the particles must be considered. It turns out that $\gamma$ never falls below 4/3, so the second core remains in hydrostatic equilibrium in spite of ionization. The complicated nature of the equation of state requires detailed calculations, whose results are usually presented in tabular form.

## 6.4  Positioning stars in the HRD

A radius of several solar radii and a mass of 0.1 $M_\odot$ brings the mean temperature, using equation (6.16), above $10^5$ K. Such a value, coupled with a mass density of order $10^{-2}\,\mathrm{gr\,cm^{-3}}$, places the object within the stellar regime. Therefore, in principle, we could place the object in the HRD and start following its track toward the zero-age main sequence. When placing it in the HRD, we plot effective temperature of the surface of the stellar-like object versus its internal luminosity, i.e., the sum of luminosity by contraction and deuterium burning. We do *not* include the accretion luminosity, which is mostly released from the inner part of the accretion disk, from small magnetic poles on the surface or from thin accretion curtains. Let us first briefly introduce and discuss the main governing factors of this evolution; they are (1) four competing timescales, (2) the three distinct sources of energy production, (3) convective versus radiative energy transport, in itself mostly controlled by (4) the source of opacity, and specifically, how opacity depends on temperature and density.

### Kelvin-Helmholtz time and accretion time

The *Kelvin-Helmholtz contraction time*

$$\tau_{\mathrm{KH}} \;=\; \frac{E_{\mathrm{pot}}}{L} \equiv \frac{G\,M^2}{R\,L} \tag{6.18}$$

$$\;=\; 3 \times 10^7 \left(\frac{M}{M_\odot}\right)^2 \left(\frac{R}{R_\odot}\right)^{-1} \left(\frac{L}{L_\odot}\right)^{-1} \quad \mathrm{yr}, \tag{6.19}$$

gives the time for an object of mass $M$ in hydrostatic equilibrium to contract from infinity to radius $R$, at an average luminosity $L$. In a slightly looser context one may view $\tau_{\mathrm{KH}}$ to be the time interval it takes the star (to be) to shrink over a factor two. Notice that the Kelvin-Helmholtz timescale gets longer as the contraction proceeds. Thus equation (6.19)

also provides an approximate measure of the total time needed for a star to contract to its main-sequence values of $M$, $R$, and $L$. For a core of 0.5 $M_\odot$ at 3 $R_\odot$ and 1 $L_\odot$, typical of the main accretion phase, this time is about 2.5 Myr.

The second timescale is the *accretion time*, which can be expressed as

$$\tau_{\text{acc}} \equiv \frac{M}{\dot{M}}, \tag{6.20}$$

where $\dot{M}$ is the mass accretion rate.

Comparing our two new timescales, we can distinguish two cases: $\tau_{\text{KH}} > \tau_{\text{acc}}$ and $\tau_{\text{KH}} < \tau_{\text{acc}}$. In the first case, the (stellar) interior cannot thermally adjust to the accretion and the luminosity will be generated by accretion shocks on the surface. As pointed out in section 6.1, this is the *proto-stellar phase*. In the second case, the accretion becomes inefficient and the star evolves towards the main-sequence on a thermal timescale with the luminosity being generated by contraction. This is called the *pre-main sequence phase*.

If we consider the expression for the mass accretion rate that follows from the simple Shu inside-out collapse model (Eq. 6.8), we can estimate the accretion time for forming a 0.5 $M_\odot$ core from an isothermal 10 K envelope to be $\sim 3 \times 10^5$ yr. This is an order of magnitude shorter than the quoted Kelvin-Helmholtz timescale. In the part of the formation process where $\tau_{\text{acc}} \ll \tau_{\text{KH}}$, the object will tend to evolve across the HRD to increasing effective temperature, essentially to reach a Hayashi track (see below) of correspondingly higher mass[2]. For most (or all, for the lowest mass stars) of the initial formation, the star will be fully convective. Whenever during this convective phase $\tau_{\text{KH}} \ll \tau_{\text{acc}}$, the star contracts and moves down the HRD following a Hayashi track. Contraction may be interrupted by deuterium burning (regardless of how $\tau_{\text{acc}}$ and $\tau_{\text{KH}}$ compare), for as long as there is sufficient fuel that the nuclear-burning timescale $\tau_{\text{N}} > \tau_{\text{KH}}$. At some point the star may develop a radiative core and deuterium may ignite in a shell around this core. This too may lead to a period of inflation, i.e., the stellar radius will temporarily increase. Later, when deuterium is depleted, the star again produces its energy solely by contraction, until finally hydrogen ignites in the core.

During the end of the main accretion phase and pre-main sequence phase mass accretion dwindles, and $\tau_{\text{KH}} < \tau_{\text{acc}}$. The Kelvin-Helmholtz contraction time lengthens as the star approaches the main sequence, which is why it is much easier to observe pre-main sequence stars that are approaching the main sequence (which are on Henyey tracks, see Sect. 6.8) than proto-stars (which are still on Hayashi tracks, see also Sect. 6.8). There are many more stars close to the main sequence. We will return to this issue in the next chapter.

The fourth characteristic timescale, the free-fall time Eq. (3.78), is for most stars much shorter than both the Kelvin-Helmholtz timescale and the accretion timescale, as well as the nuclear timescale when relevant. However, for objects that become more massive than 8 $M_\odot$, the typical Kelvin-Helmholtz timescale becomes very short, much shorter than $\tau_{\text{ff}}$. As the star

---

[2]This part of evolution is usually not shown in HRDs of star formation, but see Tout, Livio, and Bonnell (1999) for a detailed discussion including plots, e.g., their Fig. 2.

can not accrete faster than the timescale of gas infall, the star may actually reach the main sequence well before it has reached its final mass.

## 6.5 Accretion onto proto-stars

### Luminosity and radius in the main accretion phase

Gas that approaches the surface of the central object is now traveling essentially at free-fall velocity, which is considerably greater than the local sound speed. The steady rise in mass of the central object inflates this supersonic infall region, so that the cloud collapse proceeds in the usual inside-out manner. By this point, the central object is said to have entered the *main accretion phase*. At first, most of the gravitational energy goes into dissociating and ionizing hydrogen and helium, i.e. in increasing the internal energy, and only some is radiated into space.

We can estimate the maximum size that a protostar can have as well as its accretion luminosity from a simple consideration of the virial theorem, $E_{\text{th}} = -1/2\, E_{\text{grav}}$, and conservation of energy Eq. 3.48. Since the matter is initially in a cold sphere with large dimensions, both the initial internal energy and initial potential energy can be considered to be negligble. This yields

$$-\frac{1}{2}\frac{3}{5}\frac{GM^2}{R} + E_{\text{int}} + L_{\text{rad}}\,\tau = 0, \tag{6.21}$$

where $L_{\text{rad}}$ is the average luminosity over the timescale $\tau$ and $E_{\text{int}}$ is the internal energy of the gas. During the contraction, molecular hydrogen gets dissociated (the dissociation energy $E_{\text{diss}} = 4.48$ eV) and hydrogen and helium get ionized (the ionization energies $E_{\text{ion}}(\text{H}) = 13.6$ eV and $E_{\text{ion}}(\text{He}) = 79$ eV). The change in internal energy $\Delta E_{\text{int}}$ can thus be written as

$$\Delta E_{\text{int}} = \frac{XM}{m_{\text{H}}}\left[\frac{\Delta E_{\text{diss}}(\text{H}_2)}{2} + \Delta E_{\text{ion}}(\text{H})\right] + \frac{YM\Delta E_{\text{ion}}(\text{He})}{4m_{\text{H}}}. \tag{6.22}$$

$X$ and $Y$ are here the mass fractions of hydrogen and helium respectively. Ignoring $L_{\text{rad}}$ for the moment (see above) we can derive an expression for the maximum possible radius of the protostar

$$R_{\text{max}} = \frac{3}{10}\frac{GM^2}{\Delta E_{\text{int}}} = 36\,R_{\odot}\left(\frac{M}{\text{M}_{\odot}}\right). \tag{6.23}$$

Even thought this is an upper limit for the proto-stellar radius, it shows that proto-stars start out large. In the end, the immediate descendants of proto-stars, the youngest T-Tauri stars, are smaller by about an order of magnitude (as we shall see in the next chapter).

So, since we know that the true proto-stellar radii are much smaller than the derived estimate of $R_{\text{max}}$, we know that when the dissociation and ionization become complete the second term in Eq. (6.21) must become considerably smaller than the third one. Let us therefore turn

to the alternative extreme and set the third term equal to the first one. Assuming $t = \tau_{\text{acc}}$, we find

$$L_{\text{rad}} = \frac{3}{10} \frac{GM\dot{M}}{R} = \frac{3}{10} L_{\text{acc}} \qquad (6.24)$$

where we have introduced the accretion luminosity

$$L_{\text{acc}} \equiv \frac{GM\dot{M}}{R} = 314\, L_\odot \left( \frac{\dot{M}}{10^{-5}\, M_\odot \text{yr}^{-1}} \right) \left( \frac{M}{M_\odot} \right) \left( \frac{R}{R_\odot} \right)^{-1}. \qquad (6.25)$$

The accretion luminosity $L_{\text{acc}}$ is the energy per unit time released by in-falling gas that converts *all* its kinetic energy into radiation as it lands on the proto-stellar surface. Throughout the main accretion phase, regardless of the detailed time dependence of $\dot{M}$, $L_{\text{rad}} \sim L_{\text{acc}}$. Moreover, this equality holds even if the gas first strikes a circumstellar disk, then subsequently spirals onto the star. The only stipulation is that each fluid element's thermal plus kinetic energies be relatively small once it joins the proto-star. The accretion luminosity, although a product of cloud collapse, is mostly generated close to the surface of the central object. Additional sources to $L_{\text{rad}}$ come from nuclear fusion and the quasi-static contraction of the interior. However, in the proto-stellar phase these contributions to $L_{\text{acc}}$ are minor for low and intermediate masses[3] .

The radiation is able to escape from the cloud because it is gradually degraded (i.e. re-thermalised) into the infrared regime as it travels outward. Infrared photons can traverse even the large column density of dust lying between the proto-stellar surface and the outer reaches of the parent dense core. Observationally, then, proto-stars are optically invisible objects that should appear as compact sources at longer wavelengths.

Figure 6.7 sketches how the radiation diffuses outward. The figure also indicates the major physical transitions in the cloud material that is freely falling onto the proto-star. Most of the radiation is generated at the *accretion shock*. Since material further inside is settling with relatively low velocity, the shock front itself constitutes the proto-star's outer boundary. Note how the figure suggests a turbulent state for the deeper interior. Such turbulence is induced by nuclear fusion of deuterium at the center, as we will describe shortly.

The immediate post-shock temperatures at the accretion shock exceed $10^6$ K. Such hot gas emits photons in the extreme ultraviolet and soft X-ray regimes ($\lambda < 100\,\text{Å}$). The emission here is mainly in lines from highly ionized metallic species, such as Fe IX. Immediately outside of the accretion shock is a zone called the *radiative precursor*, and the material in this region is opaque to these photons. If the outer edge of the radiative precursor zone is at $R_{\text{rp}}$, the effective temperature of the radiative precursor is found approximately from $L_{\text{rad}} = 4\pi R_{\text{rp}}^2 \sigma T_{\text{rp}}^4$. Substituting for $L_{\text{rad}}$ from Eq. (6.25) and solving for the temperature, we obtain

$$T_{\text{rp}} \approx \left( \frac{3\, GM\dot{M}}{40\, \pi\sigma R_{\text{rp}}^3} \right)^{1/4} \qquad (6.26)$$

---

[3]For massive stars, significant accretion is expected to be on-going even when the star has entered the hydrogen-burning phase and thermonuclear burning is contributing strongly to $L_{\text{rad}}$.
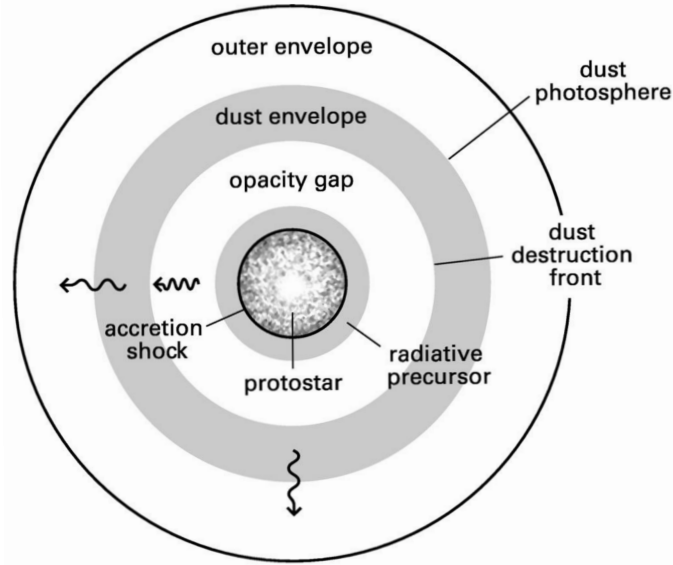
Figure 6.7: *Structure of a spherical proto-star and its infalling envelope. The relative dimensions of the outer regions have been greatly reduced in this sketch. Note the conversion of optical to infrared photons in the dust envelope. Note also the convection introduced by deuterium burning in the central, hydrostatic object. From: Stahler & Palla (2004).*

$$= 5400 \left( \frac{\dot{M}}{10^{-5} \, M_\odot \mathrm{yr}^{-1}} \right)^{1/4} \left( \frac{M}{1 \, M_\odot} \right)^{1/4} \left( \frac{R}{5 \, R_\odot} \right)^{-3/4} \quad \mathrm{K}. \qquad (6.27)$$

The quantity $T_{\mathrm{rp}}$ characterizes, at least roughly, the spectral energy distribution of the radiation field. This radiation field typically peaks in the optical spectral range. The volume outside of the radiative precursor is almost transparent to the proto-stellar radiation until it reaches the *dust envelope*. The dust envelope marks the start of the zone where the radiation temperature is 1500 K or less, i.e. where dust grains can still exist, and is opaque to the proto-stellar radiation. The location of the dust destruction front is at about 1 AU, but depends on the accretion rate.

The location where photons can start to escape freely from the dust envelope may be referred to as the *dust photosphere*, $R_{\mathrm{phot}}$, defined by

$$\tau_{\mathrm{R}} = \int_{R_{\mathrm{phot}}}^{\infty} \kappa_{\mathrm{R}}(T, r) \, \rho(r) \, dr \sim 1, \qquad (6.28)$$

where $\tau_{\mathrm{R}}$ is the Rosseland opacity. The Rosseland-opacity $\kappa_{\mathrm{R}}$ is dominated by the dust contribution. In the temperature range of interest, from about 100 to 600 K, this quantity may be represented approximately by a power law,

$$\kappa_{\mathrm{R}} \approx 4.8 \left( \frac{T}{300 \, \mathrm{K}} \right)^{0.8} \quad \mathrm{cm}^2 \, \mathrm{gr}^{-1}. \qquad (6.29)$$

One may use Figure 6.4 to verify that this is indeed roughly correct).

Material in the dust envelope is still in free-fall, therefore we may combine Eq. (6.6) and the mass continuity relation Eq. (3.15) to obtain an expression for the density profile, i.e.

$$\rho(r) = \frac{\dot{M}\, r^{-3/2}}{4\pi\sqrt{2GM}},$$

(6.30)

where $M$ is the mass of the central object (at a given time). The temperature of the dust photosphere $T_{\mathrm{phot}}$ at $R_{\mathrm{phot}}$ may be defined using

$$L_{\mathrm{phot}} = 4\pi R_{\mathrm{phot}}^2 \sigma T_{\mathrm{phot}}^4 \simeq L_{\mathrm{acc}},$$

(6.31)

where we use expression (6.25) to specify the accretion luminosity. We now have a set of two equations in the unknowns $R_{\mathrm{phot}}$ and $T_{\mathrm{phot}}$. For $\dot{M} = 10^{-5}\, M_\odot\mathrm{yr}^{-1}$ and $M = 1$, numerical solution yields $R_{\mathrm{phot}} = 2.1 \times 10^{14}$ cm (14 AU) and $T_{\mathrm{phot}} = 300$ K. So, according to Wien's law, the peak emission lies in the infrared spectral range at around $50\,\mu$m. Of course, these mean opacities and simple approximations can only give a very crude picture of what the dust photosphere is. Full frequency-dependent radiative transfer is needed to calculate the dust temperature profile as a function of optical depth.

## 6.6 Structure and evolution of the central object

Let us now shift our attention from the matter surrounding the proto-star to the structure and evolution of the central object itself. In principle, one may treat the developing proto-star as part of the solution to the large-scale hydrodynamical flow. However, such an approach is both cumbersome and prone to numerical inaccuracies. A much more natural approach is to describe the object using the stellar structure equations, i.e. the equations that are also used to describe main-sequence and post-main sequence stars. To remind the reader, these are the equations of mass continuity, hydrostatic equilibrium, state, energy transport, and energy generation:

$$\frac{\mathrm{d}M_r}{\mathrm{d}r} = 4\pi r^2 \rho$$

(6.32)

$$\frac{\mathrm{d}p}{\mathrm{d}r} = -\rho\frac{GM_r}{r^2}$$

(6.33)

$$p = \frac{\mathcal{R}\,\rho\,T}{\mu}$$

(6.34)

$$\frac{\mathrm{d}T}{\mathrm{d}r} = -L_r\left[\frac{3\kappa\,\rho}{64\pi\,r^2\,\sigma\,T^3}\right]$$

(6.35)

$$\frac{\mathrm{d}L_{\mathrm{int}}}{\mathrm{d}r} = 4\pi r^2 \rho\left[\epsilon - T\frac{\partial s}{\partial t}\right]$$

(6.36)

where $\mathcal{R}$ is the gas constant, $\mu$ the mean molecular weight, $\kappa$ the mean extinction (usually taken to be the mean Rosseland extincion $\kappa_{\mathrm{R}}$), $\epsilon$ the rate of nuclear energy generation per unit mass, and $s$ the specific entropy (i.e the entropy per units mass). The mass $M_r$ denotes the mass interior to radial distance $r$, i.e.

$$M_r = \int_0^r 4\pi r^2 \, \rho \, dr, \tag{6.37}$$

and may actually also be used as the independent variable of the system (replacing $r$). Since the proto-star's interior is highly opaque, the transport of radiation is governed by the diffusion equation (i.e. Eq. 6.35). $L_r$ is the luminosity at $r$ and may be replaced by the internal luminosity $L_{\mathrm{int}}$. The second term in the energy generation equation (6.36), $-T\partial s/\partial t$, stems from the contraction of the star.

We can solve these equations numerically by using opacities, mean molecular weights, energy generation rate and entropy tabulated as functions of $T$ and $\rho$. To close the system of equations, we need to specify the boundary conditions. For the center of the protostar, $M_r$ and $L_{\mathrm{int}}$ vanish, i.e.

$$M_r(0) = 0 \qquad \text{and} \qquad L_{\mathrm{int}}(0) = 0. \tag{6.38}$$

At the surface of the proto-star $r = R_\star$, the pressure should approach the ram pressure of the infalling gas

$$p(R_\star) = \rho \, v_{\mathrm{ff}}^2 = \frac{\dot{M} R_\star^{-3/2}}{4\pi \sqrt{2GM_\star}} \frac{2GM_\star}{R_\star} = \frac{\dot{M}}{4\pi} \left( \frac{2GM_\star}{R_\star^5} \right)^{1/2}, \tag{6.39}$$

where $\rho$ is given by Eq. (6.30) and $v_{\mathrm{ff}}$ by Eq. (6.6). Finally, the luminosity of the proto-star should be the sum of the accretion luminosity and internal luminosity

$$L_\star = L_{\mathrm{acc}} + L_{\mathrm{int}}, \tag{6.40}$$

where $L_{\mathrm{int}}$ is the luminosity at the outside boundary of the accretion shock front[4].

## Mass accretion rate and mass-radius relation

The final ingredient needed to construct proto-star models is the mass accretion rate $\dot{M}$. This quantity enters the boundary condition (6.39), and tells us, of course, how much to increase $M$ from one time step to the next. Ideally, one should take this rate directly from collapse calculations, such as the one shown in Fig. 6.3. Thus far, however, the most detailed studies of proto-star interiors have considered only constant rates, effectively treating $\dot{M}$ as a free parameter.

For guess values of the central temperature and pressure (or density) the set of stellar structure equations is solved for in the radial direction until the integration reaches $M_r = M_\star$. In

---

[4]This luminosity may be modified by back-warming from hot post-shock gas and by back-warming of accretion luminosity liberated in the hot post-shock gas emitted in the outward direction that is absorbed in the radiative precursor and re-emitted back to the shock front. See Stahler & Pally (2011), p. 328-329 for more details.

general, equations (6.39) and (6.40; recall that the accretion luminosity depends on $M_\star$ and $R_\star$ through Eq. 6.25) will not be satisfied and therefore we need to alter our guesses for the central temperature and pressure until these two conditions are met. Among other global quantities, this yields the proto-stellar radius.

The initial state, corresponding to the beginning of the main accretion phase, is somewhat arbitrary, reflecting the meager knowledge of this earliest epoch. Fortunately, the choice here has little impact on subsequent evolution. Figure 6.9 illustrates this point graphically, where five curves have different values for $R_\star$ at the initial mass, here taken to be 0.1 $M_\odot$. After the first doubling of the mass, the solutions are essentially indistinguishable. That the radius becomes a uniquely defined function of mass simply reflects that it is determined by the proto-star's internal structure set by the balance of self-gravity and thermal pressure.

## 6.7 Structure of the atmosphere

In writing down the stellar structure equation for energy transport, Eq. (6.35), we explicitly assumed that radiation is responsible for this transport. Let us and derive a simple expression for the temperature structure in the atmosphere and below. Introducing the total radiative flux $\mathcal{F} = L_r/4\pi r^2$, we may rewrite the diffusion equation as

$$\mathcal{F} = -\frac{16}{3}\frac{1}{\kappa\rho}\sigma T^3\frac{\mathrm{d}T}{\mathrm{d}r} = -\frac{4}{3}\frac{1}{\kappa\rho}\frac{\mathrm{d}(\sigma T^4)}{\mathrm{d}r} = +\frac{4}{3}\frac{\mathrm{d}(\sigma T^4)}{\mathrm{d}\tau} = \sigma T_\star^4, \tag{6.41}$$

where

$$d\tau = -\kappa\rho\,dr \tag{6.42}$$

is the optical depth interval, defined with a minus sign such that $\tau = 0$ outside of the atmosphere. The opacity $\kappa$ is in units of $\mathrm{cm^2\,gr^{-1}}$. The latter equality introduces the surface (or effective) temperature, defined at an optical depth $\tau = 2/3$. Integrating the differential equation in the third and fourth right-hand-side yields,

$$T^4 = \frac{3}{4}T_\star^4\left(\tau + \frac{2}{3}\right), \tag{6.43}$$

which describes the temperature as a function of $\tau$. One may verify that $T(\tau = 2/3) = T_\star$.

Let us also derive a simple expression for the pressure structure in the atmosphere. Hydrostatic balance requires that

$$\frac{\mathrm{d}p}{\mathrm{d}r} = -\rho\frac{GM_r}{r^2} \simeq -\rho\frac{GM_\star}{R_\star^2}, \tag{6.44}$$

where we have put $M_r \to M_\star$ and $r \to R_\star$, because the atmosphere is only a very thin layer at the surface of the proto-star. Equivalently, we may write

$$\frac{\mathrm{d}p}{\mathrm{d}\tau} = -\frac{1}{\kappa\rho}\frac{\mathrm{d}p}{\mathrm{d}r} \simeq \frac{GM_\star}{R_\star^2\,\kappa}, \tag{6.45}$$

which we may solve for a given description of the opacity $\kappa$.

When the central proto-star becomes hotter than about $2000\,\mathrm{K}$ the dust evaporates. The main opacity source is then the $H^-$ ion, for which the mean opacity in $\mathrm{cm^2\,gr^{-1}}$ is approximately given by

$$\overline{\kappa} = \kappa_\circ\, \rho^{1/2}\, T^{9/2}, \tag{6.46}$$

where $\kappa_\circ$ is a constant. So, for $H^-$ the opacity steeply rises with temperature. Substituting this behavior in Eq. (6.45) and using Eq. 6.43 and the ideal gas law Eq. (3.2) we obtain after some manipulation

$$p^{1/2}\frac{\mathrm{d}p}{\mathrm{d}\tau} = \frac{2}{3}\frac{\mathrm{d}(p^{3/2})}{\mathrm{d}\tau} = \frac{GM_\star}{R_\star^2\,\kappa_\circ}\left(\frac{\mathcal{R}}{\mu}\right)^{1/2}\frac{1}{T^4} = \frac{4}{3}\frac{GM_\star}{R_\star\,\kappa_\circ}\left(\frac{\mathcal{R}}{\mu}\right)^{1/2}\frac{1}{T_\star^4\,(\tau+2/3)}. \tag{6.47}$$

This can be integrated to give

$$p = \left(\frac{2GM_\star\,\ln{[\tau+2/3]}}{R_\star^2\,\kappa_\circ\,T_\star^4}\right)^{2/3}\left(\frac{\mathcal{R}}{\mu}\right)^{1/3}, \tag{6.48}$$

which tells us how the pressure varies in the stellar atmosphere as a function of optical depth.

## The transition point between radiative and convective energy transport

We now have expressions that describe how the temperature and pressure rise as one penetrates into the outer layers of a star. Without introduction or derivation we state that convection will develop if the temperature gradient required to drive the radiative flux becomes too steep. In the theory of Schwarzschild, the condition to prevent convection is

$$\frac{\mathrm{d}\ln p}{\mathrm{d}\ln T} > \frac{\gamma}{(\gamma-1)} = \frac{5}{2}, \tag{6.49}$$

where $\gamma$ is the adiabatic exponent, as always, and we have adopted the value $\gamma = 5/3$ appropriate for a monoatomic gas. In this expression of the Schwarzschild instability criterion we have ignored gradients in the mean molecular weight $\mu$. Using our results Eq. (6.43) and (6.48) and some of the intermediate results we may also write

$$\frac{\mathrm{d}\ln p}{\mathrm{d}\ln T} = \frac{T}{p}\frac{\mathrm{d}p/\mathrm{d}\tau}{\mathrm{d}T/\mathrm{d}\tau} = \frac{8}{3\ln{[\tau+2/3]}}. \tag{6.50}$$

It follows that convection develops once $\tau = \tau_{\mathrm{trans}} > e^{16/15} - 2/3 = 2.4$, which is just below the surface. This is for instance characteristic for the situation in the solar photosphere, that is also dominated by the opacity of $H^-$: the outer few optical depths are in radiative equilibrium, simply because at the surface photons can escape by direct flight thus making radiative transport highly efficient. At optical depths deeper than a few convection takes over, of which the solar granular pattern is evidence. Substitution of $\tau_{\mathrm{trans}}$ in equation (6.43) and (6.48) provides expressions for the pressure $p_{\mathrm{trans}}$ and temperature $T_{\mathrm{trans}}$ at the transition point from radiative to convective energy transport. One finds that $T_{\mathrm{trans}} \simeq 1.3\,T_\star$.

## 6.8  Hayashi, Henyey, and massive-star tracks

### Hayashi tracks

In the entire interior of the proto-star (inside $\tau = 2.4$) the medium is convective, so everywhere $d \ln p / d \ln T \simeq 5/2$, hence

$$\left\{ \frac{p}{T^{5/2}} \right\}_{\text{interior}} = \text{constant} = \frac{p_{\text{trans}}}{T_{\text{trans}}^{5/2}} \simeq \frac{\overline{p}}{\overline{T}^{5/2}}, \tag{6.51}$$

where $\overline{p}$ and $\overline{T}$ are mean values of the pressure and temperature in the stellar interior. We may estimate the mean pressure using Eq. 6.33 and assuming, very roughly, that $dp/dr \approx \underline{p}/R_\star$. From a global analysis we then find

$$\frac{dp}{dr} \approx \frac{\overline{p}}{R_\star} = \frac{GM_\star \rho}{r^2} \approx \frac{3}{4\pi} \frac{GM_\star^2}{R_\star^5} \quad \rightarrow \quad \overline{p} = \frac{3}{4\pi} \frac{GM_\star^2}{R_\star^4}. \tag{6.52}$$

For the mean temperature we may use the result Eq. (6.16).

Substituting into equation (6.51) for $p_{\text{trans}}$, $T_{\text{trans}}$, $\overline{p}$ and $\overline{T}$, from equations (6.48), (6.43), (6.52) and (6.16), we obtain, after some rearranging, an expression for the surface temperature, $T_\star$, of the protostar. In doing so it seems fair to ignore front constants of order unity, as it will be clear to the reader that the result should be treated as approximate given all the assumptions that have entered the derivation. We arrive at

$$T_\star \simeq \left( G^{13} \, (\mathcal{R}/\mu)^{17} \, \kappa_\circ^{-4} \, M_\star^7 \, R_\star \right)^{1/31} \tag{6.53}$$

It follows that the internal luminosity, $L_\star \simeq 4\pi R_\star^2 \sigma T_\star^4$, is

$$L_\star \simeq 4\pi\sigma \left( G^{52} \, (\mathcal{R}/\mu)^{68} \, \kappa_\circ^{-16} \, M_\star^{28} \, R_\star^{66} \right)^{1/31} \tag{6.54}$$

Eliminating $R_\star$ between the above two equation, we obtain an expression for the luminosity in terms of the star's mass and surface temperature

$$L_\star \propto M_\star^{-14} \, T_\star^{66} \qquad \text{hence} \qquad \left. \frac{d \ln L_\star}{d \ln T_\star} \right|_{M_\star} = 66. \tag{6.55}$$

Rearranging yields for the temperature in terms of the star's mass and luminosity

$$T_\star \propto M_\star^{0.212} \, L_\star^{0.015} \qquad \text{hence} \qquad \left. \frac{d \ln T_\star}{d \ln M_\star} \right|_{L_\star} = 0.212. \tag{6.56}$$

So let us now return to our discussion of the evolution of the star. The proto-star's luminosity is given by Eq. 6.40. Over time, the mass-accretion rate diminishes and the luminosity by

contraction becomes more and more important. That luminosity starts to define the position and evolution of the protostar on the Hertzsprung-Russell (HR) diagram (see Figure 6.8). From equation 6.55, it follows that the tracks for a star of given mass are almost vertical on the HR diagram. From equation 6.56 it follows that tracks for more massive stars are hotter (but only slightly hotter) than those for lower-mass stars.

Since the luminosity is supplied more and more by gravitational contraction alone, the evolution is in the direction of decreasing $R_\star$, and hence of decreasing $L_\star$. Thus stars in this phase evolve downwards on the HR diagram. This convective contraction part of the evolution track is called the *Hayashi track*. Realize that it is essentially the opacity, Eq. 6.46, that is controlling this behavior.

## Henyey tracks

As the proto-star continues to contract and heat up (its interior), bound-free and free-free processes become the dominant source of opacity, given by

$$\overline{\kappa} = \kappa_1 \, \rho \, T^{-7/2}, \tag{6.57}$$

where $\kappa_1$ is a normalization constant and the unit is again $\mathrm{cm^2\,gr^{-1}}$. As the opacity is very much in control of the way the star evolves through the HR-diagram one may expect a different behavior as during the Hayashi phase, where the opacity is proportional to $\rho^{1/2}\,T^{9/2}$ (see Eq. 6.46). The increasing temperature now leads to a reduction of the opacity. The combination of the decreasing luminosity and the increasing (internal) temperature eventually cause the star to switch back from energy transport by convection to radiation.

We again resort to a dimensional analysis to establish how the star will evolve through the HR-diagram. In radiative energy transport the luminosity is given by

$$L_\star \simeq \frac{R_\star \, \sigma \overline{T}^4}{\overline{\kappa}\rho}, \tag{6.58}$$

which follows directly from Eq. (6.35). Equating this to $L_\star = R_\star^2 \, \sigma T_\star^4$ yields

$$T_\star \propto \overline{T} \, (\overline{\kappa}\rho \, R_\star)^{-1/4} = \overline{T} \, \overline{\tau}^{-1/4}, \tag{6.59}$$

where $\overline{\tau}$ is the opacity from the center to the surface. Using Eqs. (6.57) and (6.16) for the opacity and mean internal temperature, we obtain the optical depth in the form

$$\overline{\tau} \simeq \overline{\kappa}\rho \, R_\star \sim \kappa_1 \left( \frac{G\,\mu}{\mathcal{R}} \right)^{-7/2} (M_\star R_\star)^{-3/2} \tag{6.60}$$

Substituting from Eqs. (6.16) and (6.60), we obtain for the surface temperature

$$T_\star \simeq \kappa_1^{-1/4} \left( \frac{G\,\mu}{\mathcal{R}} \right)^{15/8} M_\star^{11/8} \, R_\star^{-5/8}. \tag{6.61}$$
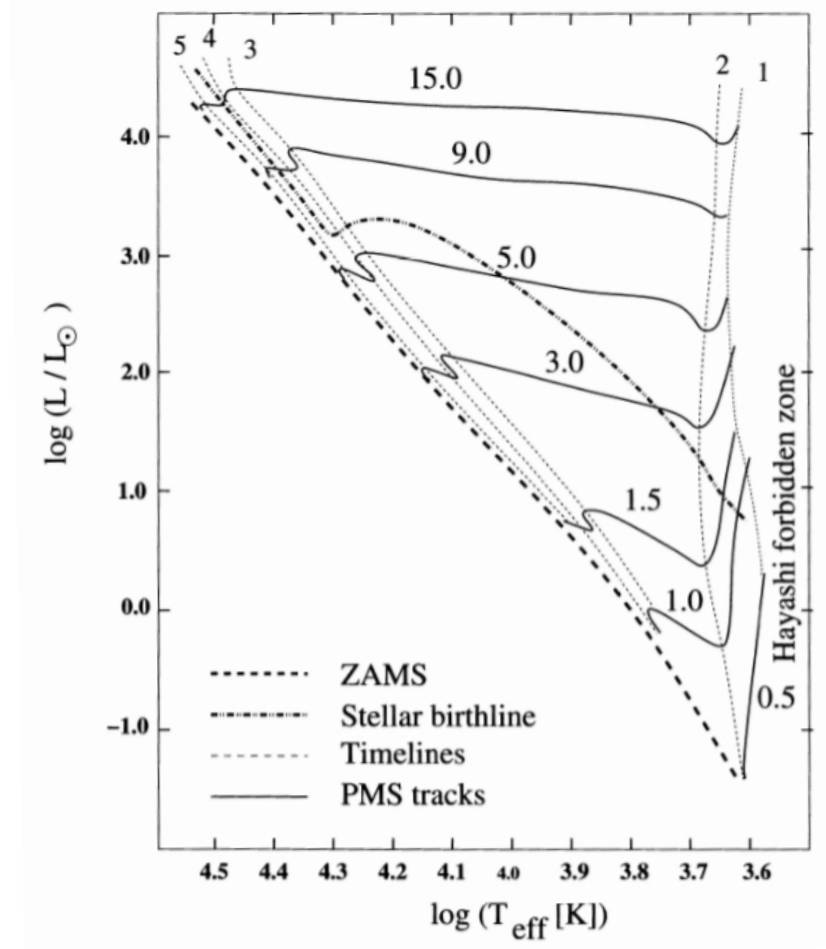
Figure 6.8: *Evolutionary path in the Hertzsprung-Russell (HR) diagram. Thin dashed lines labelled 1, 2, 3, 4, and 5 indicate timelines. Timeline 1 is the beginning of the convective Hayashi track, timelines 2-5 cover the radiative Henyey track of the HR-diagram. The thick dahsed line is the zero-age main sequence; the thick dash-dotted line the stellar birthline for an accretion rate of $10^{-5}$ $M_\odot \mathrm{yr}^{-1}$. From: Schulz 2004.*

It follows that the luminosity, using $L_\star = R_\star^2 \, \sigma T_\star^4$, is

$$L_\star \simeq \sigma \kappa_1^{-1} \left( \frac{G\mu}{\mathcal{R}} \right)^{15/2} M_\star^{11/2} \, R_\star^{-1/2}. \tag{6.62}$$

Eliminating $R_\star$ between the above two equations, we obtain

$$L_\star \propto M_\star^{22/5} \, T_\star^{4/5} \quad \text{hence} \quad \left. \frac{\mathrm{d}\ln L_\star}{\mathrm{d}\ln T_\star} \right|_{L_\star} = 0.8 \quad \text{and} \quad \left. \frac{\mathrm{d}\ln L_\star}{\mathrm{d}\ln M_\star} \right|_{L_\star} = 4.4. \tag{6.63}$$

It follows that the tracks are diagonal (from bottom right to top left) on the HR-diagram. Also,

tracks for more massive stars are (well) above those for less massive stars. This too is clearly visible in Fig. 6.8.

As the luminosity is supplied by gravitational contraction, the evolution is in the direction of decreasing $R_\star$, and hence in the direction of increasing $T_\star$ and increasing $L_\star$ (see Eqs. 6.61 and 6.62). In other words, in this radiative contraction phase the star evolves upwards and to the left on the HR diagram; it is sometimes referred to as the *Henyey track*.

## Massive-star tracks

For massive proto-stars, the opacity switches to Thomson scattering by free electrons. This opacity starts to dominate when the medium becomes fully ionized and is given by the simple expression

$$\kappa = \kappa_2, \tag{6.64}$$

where $\kappa_2$ is a normalization constant and again $\kappa$ is in cm$^2$ gr$^{-1}$. The mean optical depth from center to surface is now

$$\overline{\tau} = \kappa \, \overline{\rho} \, R_\star \simeq \frac{\kappa_2 \, M_\star}{R_\star^2}. \tag{6.65}$$

Here too, the opacity is low enough for radiative transport of energy, and we may use Eqs. (6.16) and (6.65) in Eq. (6.59) to obtain an expression for the surface temperature of a massive star on its Henyey track

$$T_\star \simeq \overline{T} \, \overline{\tau}^{-1/4} \simeq \left( \frac{G \, \mu}{\mathcal{R}} \right) \kappa_2^{-1/4} \, M_\star^{3/4} \, R_\star^{-1/2}. \tag{6.66}$$

The luminosity again follows using $L_\star = R_\star^2 \, \sigma T_\star^4$, resulting in

$$L_\star \simeq \sigma \left( \frac{G \, \mu}{\mathcal{R}} \right)^4 \kappa_2^{-1} \, M_\star^3. \tag{6.67}$$

This final equation shows that Henyey tracks for massive stars should be approximately horizontal in the HR-diagram. In other words, for a star of given mass, the star evolves onto the main sequence at approximately constant luminosiy (see Fig. 6.8).

Since the luminosity is supplied by gravitational contraction, the evolution is in the direction of decreasing $R_\star$, and hence (from Eq. 6.66) in the direction of increasing $T_\star$. Consequently, we see that massive stars simply evolve horizontally to the left on the HR diagram.

## 6.9 Deuterium burning and hydrogen ignition

The first nuclear fuel to ignite is a small admixture of the hydrogen isotope deuterium (D or $^2$H). Deuterium is the product of primordial nucleosynthesis, produced during the universal
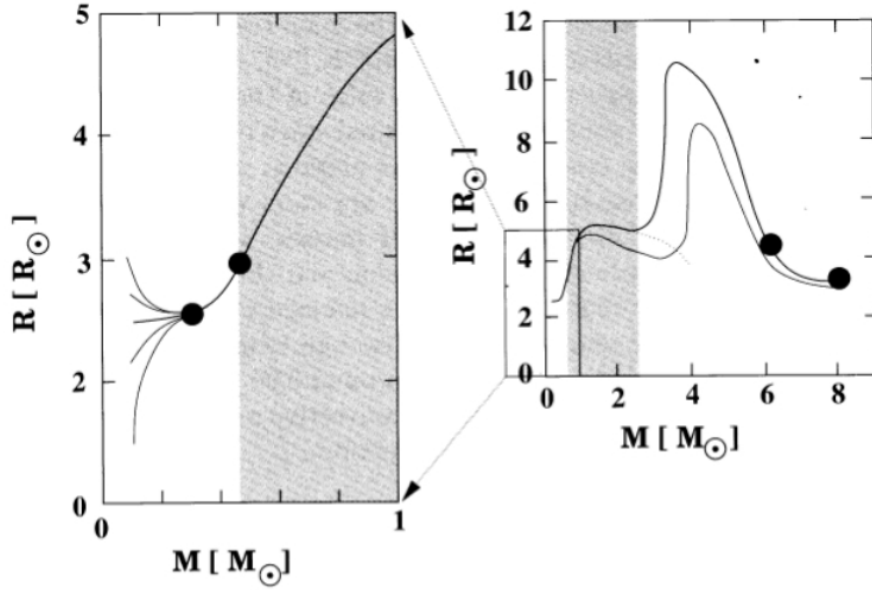
Figure 6.9: *Evolution of stellar radius with mass during the assembly of what ultimately will be a 8 $M_\odot$ star using a constant accretion rate of $10^{-5}$ $M_\odot$/yr. The horizontal axis is essentially also a linear time axis as $M(t) = M_o + \dot{M} \cdot t$, with $M_o$ the initial mass of the second core. The grey area indicates where the protostars are fully convective; black dots mark (from left to right) the onset of deuterium burning; the point where stars are fully convective; the appearance of a central convection zone because of the ignition of hydrogen in the core (H-burning starts at $\sim 5$ $M_\odot$), and the arrival at the ZAMS. The thin lines in the left panel denote different values for $R_{\text{star}}$ at the initial mass. Note that by the time the stellar mass has doubled, the curves are nearly identical. The thin line in the right panel also illustrates the use of different boundary conditions. From: Schulz 2004.*

cooling and expansion, minutes after the Big Bang. Once incorporated into stars, the isotope is destroyed by fusion with protons following the reaction

$$^2\text{H} + {}^1\text{H} \to {}^3\text{He} + \gamma. \tag{6.68}$$

This reaction is exothermic, releasing $\Delta E_D = 5.5$ MeV. As in many nuclear processes, the associated rate is highly sensitive to the local temperature. Deuterium fusion first becomes appreciable near $10^6$ K. Close to this temperature, the energy generation term in equation (6.36) may be approximated by

$$\epsilon_D = [\text{D/H}] \, \epsilon_o \left( \frac{\rho}{1 \, \text{gr cm}^{-3}} \right) \left( \frac{T}{10^6 \, \text{K}} \right)^{11.8}, \tag{6.69}$$

where $\epsilon_o = 4.19 \times 10^7$ erg gr$^{-1}$ sec$^{-1}$, and [D/H] is the interstellar number density of deuterium relative to hydrogen. Analysis of the absorption lines of stars lying behind diffuse clouds gives a mean [D/H] of $2 \times 10^{-5}$, with the apparent variation being about a factor of two.

LOW MASS                    INTERMEDIATE MASS

D-core burning
in convective core

core deprived of
freshly-accreted D
and D-core burning
dwindles

radiative barrier develops

fully convective
(hence efficient mixing)

D-core burning seized,
and core becomes
fully radiative

steady-state
D-core burning

D-shell burning around
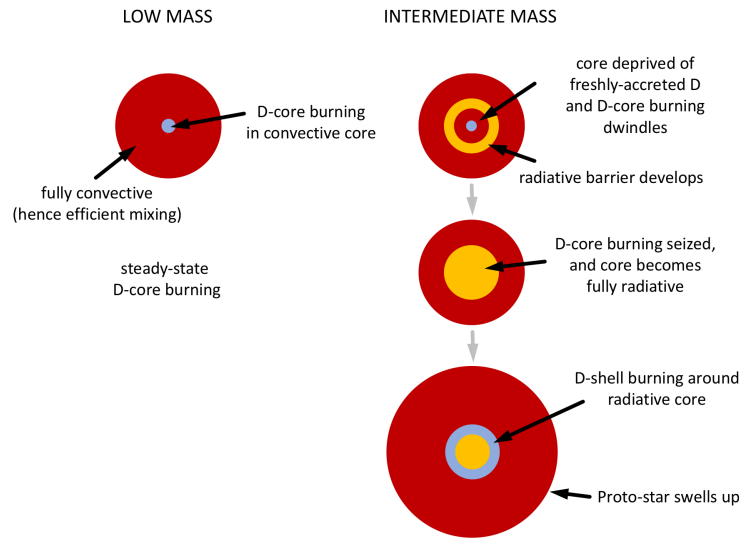radiative core

Proto-star swells up

Figure 6.10: *The four stages of deuterium burning in proto-stars. Active burning begins at the center. For low-mass stars this is the only stage. Sources that can accrete to become intermediate mass stars develop a radiative barrier. Once the core is depleted from deuterium, it too becomes radiatively stable. Fuel later re-ignites in a thick shell around the core which leads to an inflation of the proto-star. Note that the radii of the four states are not to scale. Blue regions denote D-burning; orange shows radiative zone, and red shows convective zones.*

Once deuterium burning ignites near the center of a low-mass proto-star, it produces too much luminosity to be transported radiatively through the highly opaque interior. So, convection takes over which quickly spreads out from the center. Soon, the whole interior is convectively unstable, apart from a thin, outer settling region of negligible mass. The first black dot in Fig. 6.9 marks the onset of deuterium burning and the second black dot the point where the star becomes fully convective. Core deuterium burning is the cause of the swelling of the stellar radius between the location of these two black dots, though the degree of swelling depends somewhat on the assumed accretion rate. From equation (6.16) we know that the interior temperature scales with $M_\star/R_\star$. Because the energy production rate $\epsilon_D$ is so temperature-sensitive, a further rise of the central temperature would lead to a dramatic increase of the amount of energy to be pumped through the interior layers, causing a swelling of the star. As $T_\star \propto M_\star/R_\star$ the central temperature will go down. In this manner, *deuterium acts as a thermostat*, trying to maintain the central temperature close to the ignition value of $1 \times 10^6$ K and to have $R_\star$ increase proportional to $M_\star$.

The thermostat is only effective if the proto-star has a steady supply of the nuclear fuel. In addition, the temperature-sensitivity of $\epsilon_D$ implies that active burning is confined to the central region. The original deuterium residing here is quickly consumed after ignition. While additional fuel continually lands on the surface through infall, it cannot reach the deep interior as long as the proto-star is radiatively stable. Once the interior is fully convective, however,
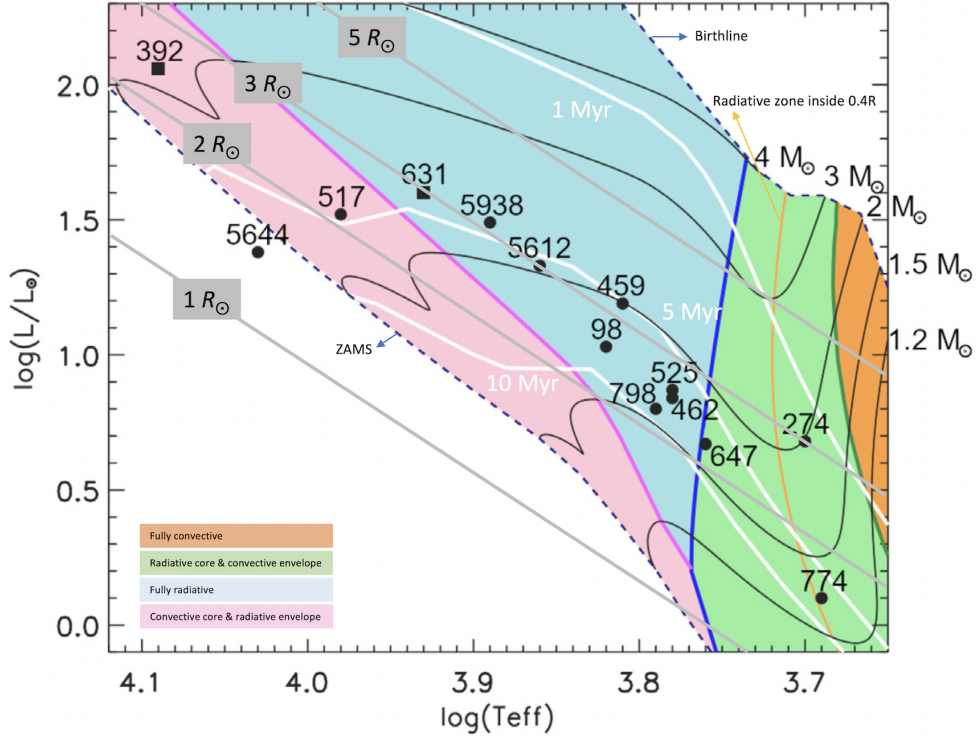
Figure 6.11: *Evolutionary path in the HRD, similar to Fig. 6.8, together with observations of 14 young stellar objects in NGC 2264 (Cody et al. 2014). The white lines correspond to the 1, 5, and 10 Myr isochrones. The black lines are the PMS evolution tracks for specific masses (see labels in plot). The birthline and ZAMS are represented by dashed lines. In the orange region, stars are fully convective; in the green region stars present a radiative core and a convective envelope, the blue region corresponds to fully radiative systems, and the pink region contains stars with a convective core and a radiative envelope. The orange line is the location where the radius of the onset of convection is 40% of the stellar radius. In grey are lines of constant radius. Notice that the onset of D-shell burning, close to the orange line, leads to an inflation of the star, as depicted in Fig. 6.10. Adapted from: Moura et al. 2020 and Villebrun et al. 2019.*

turbulent eddies drag this deuterium toward the central furnace. Because the associated transport time is relatively brief, the consumption rate approaches that supplied from infall. Thus, the total luminosity $L_D$ generated by deuterium is close to its steady-state value

$$L_D = \int_0^{R_\star} 4\pi \, r^2 \, \epsilon_D \, \rho \, dr \approx \dot{M}_{acc} \, \delta \simeq 12 \left( \frac{\dot{M}_{acc}}{10^{-5} \, M_\odot \mathrm{yr}^{-1}} \right) L_\odot. \qquad (6.70)$$

Here, $\delta$ is the energy available in deuterium per gram of interstellar gas

$$\delta = \frac{[\mathrm{D/H}] \, X \, \Delta E_D}{m_H}, \qquad (6.71)$$

where $X$ is the hydrogen mass fraction. The steady-state $L_D$ is $12 \, L_\odot$ for an accretion rate of $1 \times 10^{-5} \, M_\odot \mathrm{yr}^{-1}$. For comparison, the accretion luminosity for this accretion rate is

$60\ L_\odot$. In summary, the swelling of the proto-star actually stems from two distinct sources – backheating from the accretion shock[5], and internal heating from deuterium burning, which gives a smaller, but still important contribution.

We now continue our discussion of the construct of stars, by solving the four stellar structure equations, regarding $\dot{M}$ as a freely specifiable function. For the typical accretion rates that we consider, computations show that $M_\star$ cannot increase far above $1\ M_\odot$ before the proto-star's interior, which was fully convective from deuterium burning, reverts to a radiatively stable state. It does so because once core deuterium burning is established (and gets exhausted in the central regions) the further increase in mass occurs at roughly constant radius (see the flat part of the curve in the grey zone in Fig. 6.9), hence the internal temperature rises (see Eq.6.27) and the average opacity, as discussed in Sect. 6.8, decreases. This makes it easier for the interior luminosity to reach the surface by means of radiative transport.

The switch back to radiative transport for models with a constant mass accretion rate of $1 \times 10^{-5}\ M_\odot \mathrm{yr}^{-1}$ occurs when $M_\star$ reaches $2.38\ M_\odot$ (see also Fig. 6.9). A radiative zone establishes itself in the inner $\sim$30 percent by mass of the envelope surrounding the D-burning core and acts as a barrier that will quickly alter the proto-star's thermal structure. Once the barrier is established, the residual deuterium inside it is consumed rapidly, and convection disappears throughout the interior volume. As long as cloud collapse and infall persists, deuterium accumulates in a thick convective mantle outside the original radiative barrier. Soon, even the base of the deuterium mantle reaches $10^6$ K. The fuel ignites and causes a swelling of the star (see the steep increase in $R_\star$ to the right of the grey zone in Fig. 6.9). Once again, fresh deuterium arrives through infall and is quickly advected to the burning layer. This deuterium shell burning also approaches the steady-state level given by equation (6.70).

Figure 6.10 summarizes pictorially the stages of deuterium fusion in proto-stars. The shift in the site of active burning from the depleted central region to an overlying shell is reminiscent of the progression of hydrogen fusion during post-main-sequence evolution. In both cases, the establishment of the shell source is accompanied by major structural changes. In figure 6.9 (right-side panel) the radiative barrier establishes itself just to the right-side edge of the shaded area. Deuterium shell burning sets in somewhat later and the star reacts with a strong increase of its radius.

Notice how the first rise in $R_\star(M_\star)$ in Fig. 6.9 from central deuterium burning occupies a relatively narrow mass range. Our knowledge of the initial mass function tells us that the majority of proto-stars are actually in this interval, while only a small fraction ever attain the second swelling due to shell burning.

### Contraction and hydrogen ignition

If we continue adding mass to the proto-star, both the convection and swelling gradually disappear. Now that the star is almost fully radiatively stable, the internal luminosity is (see also

---

[5]The backheating provides about $(3/4)\ L_\mathrm{acc}$ in luminosity.

Eq. 6.62)

$$L_{\text{int}} \approx 1 \left( \frac{M_\star}{1\,M_\odot} \right)^{11/2} \left( \frac{R_\star}{1\,R_\star} \right)^{-1/2} L_\star. \tag{6.72}$$

Actually, this approximate relation is valid for *any* radiative star, regardless of its evolutionary state. Indeed, we have been able to assign the numerical coefficient from the fact that the Sun itself is stable against convection throughout most of its volume. The equation indicates that the proto-star's interior luminosity soon exceeds that produced by steady-state deuterium shell burning (Eq. 6.70). For $M_\star > 3 M_\odot$, the interior contribution amounts to several $100\,L_\odot$ and even dominates $L_{\text{acc}}$. Between 5 and 6 $M_\odot$, the luminosity surpasses $10^5\,L_\odot$. The energy source that accounts for this rise in luminosity is simply gravitational attraction of the bulk interior. This energy source has been available from the earliest times, but does not truly come into play until after the cessation of central deuterium burning. From then on, the influence of self-gravity keeps building until it becomes paramount. Figure 6.9 shows how even the large swelling due to shell ignition is soon reversed, and the star begins a rapid, overall contraction.

Consider the observational and evolutionary status of the star during this period. Our assumption of continuing infall means that the brightening intermediate-mass (i.e. few to 10 $M_\odot$) object is obscured at optical wavelengths by its dusty envelope. However, the fact that most of the luminosity now stems from internal gravitational contraction violates the definition of proto-star offered earlier. However, it can also not be claimed that the object is in the pre-main-sequence phase, where any mass addition occurs at a much slower rate than the stellar contraction. We will refer to these objects as *intermediate-mass proto-stars*, while recognizing their rather unique, hybrid nature.

It is important to realize that the accreting star has not entered a state of dynamical collapse like that which terminated the first core. The timescale for the radius to decrease significally is therefore much longer than $\tau_{\text{ff}}$, and is set by the magnitude of the radiative losses (so essentially $\tau_{\text{KH}}$). The decline of $R_\star$ (see downhill slope at the right in Fig. 6.9) implies an even faster rise of the interior temperature. Once the central value of the temperature surpasses $1 \times 10^7$ K, the protostar begins to fuse ordinary hydrogen. Hydrogen burning *in proto-stars* commences when the total mass reaches about 5 $M_\odot$. Initially, pairs of protons begin to combine

$$^1\text{H} + {}^1\text{H} \rightarrow {}^2\text{H} + e^+ + \gamma. \tag{6.73}$$

The deuterium produced here almost immediately fuses with another proton to create $^3$He, through the much faster reaction of Eq. 6.68. The subsequent conversion of $^3$He to $^4$He proceeds along a number of different pathways, that in the end imply a net reaction

$$4\,^1\text{H} \rightarrow {}^4\text{He} + 2\,e^+ + 2\,\gamma, \tag{6.74}$$

for which the energy output is 26.7 MeV. Once the proto-star's mass is about 6 $M_\odot$ and its central temperature is $2 \times 10^7$ K, the so-calles CN cycle predominates. It is clear from Fig. 6.9 that the contraction begins to slow at just this point. While Eq. (6.72) still governs the *total* interior luminosity, the contribution from nuclear burning grows at the expense of that from gravitational contraction. Soon a central convection zone appears (third black dot in Fig. 6.9).

This zone extends to $M_r = 1.3M_\odot$ by the time contraction truly halts, at a proto-stellar mass near 8 $M_\odot$. The object is now more accurately described as an accreting main-sequence star. Infall makes only a minor contribution to the total luminosity of $3.5 \times 10^3 L_\odot$, and continues to provide an obscuring screen for the emitted optical and ultraviolet photons.

### Exercise 6.1

Show that equations (3.3), (3.59), and (3.78) lead to a mass-accretion rate $\dot{M} \propto a^3/G$, when assuming $\dot{M} \propto M/\tau_{\text{ff}}$

### Exercise 6.2

We consider a homogeneous spherical cloud of radius $R = 0.2\,\text{lyr}$ and mass $M = 2.5$ $M_\odot$ consisting of molecular gas with mean molecular weight $\mu = 2.3$. The gas is interspersed with a small component of dust which dominates the opacity. The temperature is $10\,\text{K}$ and the gas particle density is $n = 5 \times 10^4$ cm$^{-3}$. The dust-to-gas ratio by mass $f_{\text{dg}} = 0.01$.

a) Read off the dust absorption coefficient $\kappa$ in cm$^2$ gr from figure 6.4.

b) Compute the optical depth for a beam passing through the center of the cloud.

c) The cloud collapses homogeneously. For what radius in AU does the cloud become optically thick for a beam passing through its center?

### Exercise 6.3

In this problem you will calculate the approximate evolution of a proto-star through the infrared HR diagram. Assume that the proto-stellar core accretes mass at a rate

$$\dot{M} = \dot{M}_\circ \left(1 - \frac{t}{\tau}\right), \qquad (6.75)$$

where $\dot{M}_\circ = 2 \times 10^{-6}\ M_\odot \text{yr}^{-1}$ is the initial accretion rate, $t$ is the time, and $\tau = 0.5\,\text{Myr}$ is the time when accretion stops.

a) Calculate the mass as a function of time and the final mass in $M_\odot$.

b) Obtain a formula for the accretion luminosity of the core as a function of time. Assume that the core radius is constant at $3R_\odot$.

c) Produce a table of $L_{\text{acc}}$, $M$, and $\dot{M}$ at 19 equal time steps starting at $0.05\tau$ and ending at $0.95\tau$.

d) At each time, calculate the approximate mean photospheric temperature and plot the evolution in the HR diagram.

e) At each time, find the approximate wavelength at maximum intensity in the observed continuous spectrum.

**Exercise 6.4**

We observe a proto-stellar object at a distance of 80 parsecs. With ALMA we can spatially resolve the dust envelope in which the protostar is embedded and find an angular diameter of 0.3 arcsec. Our goal is to make a simple model that describes the size of the dust envelope, and to use this model to constrain the mass accretion rate $\dot{M}_{\mathrm{acc}}$.

a) Compute the radius $R_{\mathrm{phot}}$ of the dust photosphere in AU.

b) We assume that the dust extinction coefficient is constant and equal to $\kappa_{\circ} = 3.0\,\mathrm{cm^2\,gr^{-1}}$. In the dust envelope and outside of $R_{\mathrm{phot}}$ the material is in free-fall, hence the density profile is given by Eq. 6.30. Use Eq. 6.28 to derive an expression for $R_{\mathrm{phot}}$ in terms of $M$ and $\dot{M}_{\mathrm{acc}}$.

c) We adopt a mass $M = 1.5 M_{\odot}$ for our source. Compute $\dot{M}_{\mathrm{acc}}$ in $M_{\odot}\mathrm{yr}^{-1}$. Recall that $1 M_{\odot}\mathrm{yr}^{-1} = 6.303 \times 10^{25}\,\mathrm{gr\,s^{-1}}$.

# 7

# Observations of star formation

The previously sketched picture of a spherical inside-out collapse cannot explain a series of observational phenomena that accompany star formation. These are for example the presence of proto-planetary disks, jets and outflows, and planets. Other processes than gravity need to be at work to 'prevent' the infall of being truly radial. While magnetic fields or rotation will not prevent the collapse, they will change the way matter flows and accretes.

## 7.1 Rotating collapse and the origin of the disk

We now briefly consider the impact of rotation on the proto-stellar collapse. Since the cloud remains hydrodynamically stable for a long time, we can assume that it rotates as a solid body. Also, we assume that the cloud rotates slowly, so that the centrifugal force is initially very small and the early stages of collapse are almost radial. In this limit, we can distinguish between an outer envelope which stays almost spherical and an inner region that gets distorted due to rotation and forms an equatorial accretion disk (Fig. 7.1).

Depending on the specific angular momentum $j$ of a gas parcel, it will either be able to accrete onto the proto-stellar core (when it has very low specific angular momentum) or accrete onto the equatorial disk at some radius corresponding to its specific angular momentum. The maximum possible distance from the star at which such a gas parcel can end up is called the centrifugal radius $r_{cf}$. This is also the quantity limiting the size of the accretion disk. We will derive this quantity from simple principles.

If we study the trajectory of a fluid element during the infall, we can assume that the proto-star + disk constitute a point source and that only gravity and rotation play a role. Under these circumstances, the fluid element will have a bound elliptical orbit[1]. However, since the inital energy is small compared to the energy it has when it reaches the equatorial plane, we can

---

[1]The conic section corresponding to a negative energy – hence an ellipse or circle; unbound orbits have conic sections corresponding to a parabola or hyperbola.
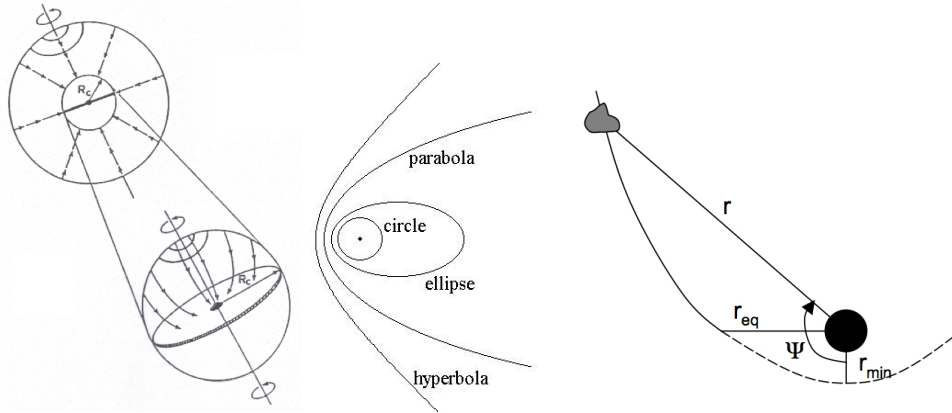
Figure 7.1: *Sketches of the geometry of the collapse of a cloud including rotation. Left panel: in the scenario of inside-out collapse the outer envelope remains spherical, while the inner region is affected by rotation. The radius dividing these two zones is the centrifugal radius $r_{cf}$ (here denoted as $R_c$). Inside the centrifugal radius, an equatorial accretion disk forms. Middle panel: though particles from the cloud fall in on bound, thus elliptical orbits, their trajectories can be well represented by parabolic orbits, corresponding to zero energies because their initial energies are small compared to that once they approach the forming disk. Right panel: a fluid element with polar coordinates $(r, \Psi)$ falls in on a parabolic orbit and hits the equatorial disk at a radius $r_{eq}$. In the absence of the disk, the fluid element would have reached the smaller distance $r_{min}$ before swinging back out again.*

approximate it reasonably well by a parabola[2]

$$r = \frac{r_{\mathrm{eq}}}{1 + \cos \Psi}, \qquad (7.1)$$

where the coordinates and quantities are introduced in the right-hand-side panel in Fig. 7.1. If the fluid element would reach $r_{\mathrm{min}}$, the maximum orbital velocity would be the free-fall velocity (see Eq. 6.6)

$$v_{\mathrm{max}}^2 = \frac{2GM_*}{r_{\mathrm{min}}}, \qquad (7.2)$$

where $M_*$ is the mass of the proto-star + disk. However, the fluid element will never reach that part of the trajectory since it collides with the forming disk at $\Psi = 90^\circ$. The specific angular momentum can be written as

$$j^2 = r_{\mathrm{min}}^2 \, v_{\mathrm{max}}^2 = 2GM_* \, r_{\mathrm{min}}. \qquad (7.3)$$

As the fluid element reaches $r_{\mathrm{min}}$ at $\Psi = 0$, Eq.(7.1) implies that $r_{\mathrm{min}} = r_{\mathrm{eq}}/2$. Therefore, we can write the radius at which the fluid element reaches the plane as

$$r_{\mathrm{eq}} = \frac{j^2}{GM_\star}. \qquad (7.4)$$

---

[2]The conic section corresponding to zero-energy – hence an unbound orbit.

Now we can express the specific angular momentum in terms of the initial quantities, i.e. the angular velocity $\Omega_0$ and the inclination of the orbital plane with respect to the rotation axis $\theta_0$ ($\theta_0 = 90^\circ$ is the midplane)

$$j = r^2 \, \Omega_\circ \, \sin \theta_\circ = a^2 \, t^2 \, \Omega_\circ \, \sin \theta_\circ, \tag{7.5}$$

where $r$ is the distance from the center at which the fluid element starts its collapse. We may identify this radius with the position of the rarefaction wave at time $t$. It moves at the sound speed $a$ and hence $r = at$. The angular momentum is larger for higher values of $\theta_\circ$ and has a maximum at $\theta_\circ = 90^\circ$. Inserting this into Eq.(7.4)

$$r_{\rm eq} = \frac{j^2}{GM_\star} = \frac{r^4 \, \Omega_\circ^2 \, \sin^2 \theta_\circ}{GM_\star} = \frac{a^4 \, \Omega_\circ^2 \, \sin^2 \theta_\circ}{GM_\star} t^4 = \frac{a \, \Omega_\circ^2 \, \sin^2 \theta_\circ}{m_\circ} t^3, \tag{7.6}$$

where we have used Eq. (6.4) and $m(x) = m_\circ$ in writing down the last right-hand-side. The maximum value for $r_{\rm eq}$ is

$$r_{\rm eq} = \frac{r^4 \, \Omega_\circ^2}{GM_\star} = \frac{a \, \Omega_\circ^2}{m_\circ} t^3. \tag{7.7}$$

Initially, the low $j$ shells will collapse (because at low $t$, $r$ is still small), and matter falls directly onto the proto-star because $r_{\rm eq} < R_\star$. Later, the matter passes the star, *collides with material from the other side* and forms a disk. This sets the value of the centrifugal radius $r_{\rm cf}$. This collision process, as a result of the symmetry of the infall, is essential in disk formation and actually is the origin of proto-planetary disks. For $\theta_\circ = 90^\circ$, $r_{\rm eq}$ is the centrifugal radius of the disk

$$r_{\rm cf} \approx 7.33 \, {\rm AU} \left( \frac{a}{0.35 \, {\rm km/s}} \right) \left( \frac{\Omega_\circ}{10^{-14} \, {\rm s}^{-1}} \right)^2 \left( \frac{t}{10^5 \, {\rm yr}} \right)^3. \tag{7.8}$$

For the latter expression, we assumed that $m_0 \sim 1$ (see Sect. 6.2). The value $\Omega = 10^{-14} \, {\rm s}^{-1}$ is close to typical observed values of a few times $10^{-14} - 10^{-13} \, {\rm s}^{-1}$. This gives us also an estimate of the initial size of the accretion disk. Assuming a constant mass-accretion rate (which we have done by adopting Eq. 6.4), the disk-radius growth is quite fast and implies that the process must occur essentially during the earliest stages of the proto-star evolution, where it remains largely hidden to an outside observer, except at the longest wavelengths. Later the size of the disk also grows due to e.g. angular momentum transport, so-called viscous spreading of the disk.

The first stages of proto-star evolution mentioned above correspond to the Class 0 and Class I stages, discussed in the next section. The observational study of the disk formation phase in then best performed at millimeter wavelengths. The first such studies revealed that disks indeed occur in Class-0 and Class-I sources, with typical masses of the order of a few percent of a solar mass. No significant mass increases are found between these two phases, suggesting that *(i)* the disks are mostly formed during the Class-0 phase, and *(ii)* the lower mass-accretion rates from the envelope during the Class-I phase are well matched with matter transferred from the inner edge of the disk onto the proto-star.
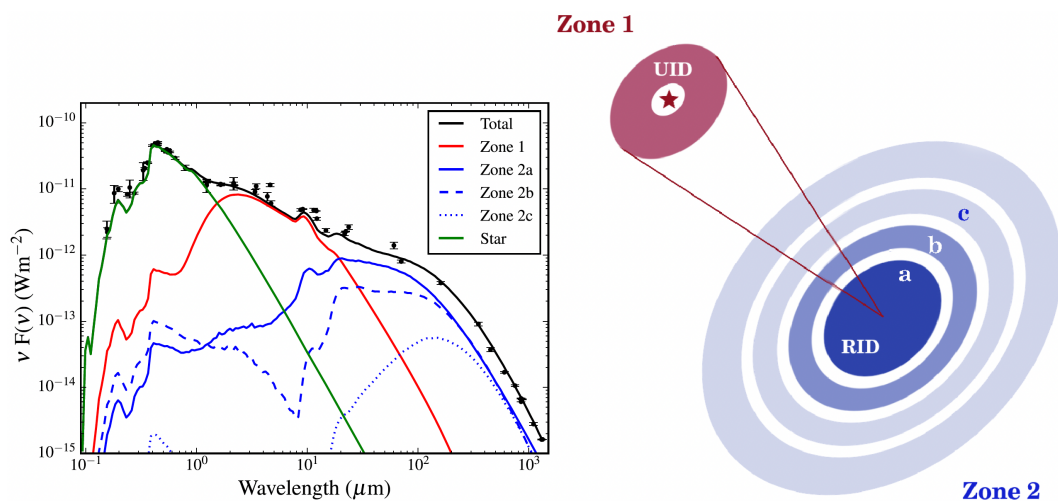
Figure 7.2: *A fit to the SED of the Herbig A1 Ve star HD 163296 (9 200 K and 33 $L_\odot$) over four orders of wavelength, from the UV to the mm, to illustrate which parts of the star-disk system contribute to the observed flux. The contribution of the central star is shown in green. The (in their observations spatially) unresolved inner disk (UID) is heated most efficiently by the central star and dominates the near-IR flux (red curve). Increasingly outward parts of the disk dominate the flux at increasingly longer wavelengths. The physical dimensions of the MCMax3D disk model zones 2a, 2b, and 2c are 1.70−63, 63−105, 105−240 AU. From: Muro-Arena et al. (2018).*

## 7.2  The spectral energy distribution of star-forming systems

We will now take a more observational approach and discuss how the progressive phases of star formation can be empirically identified. Observing the star-forming regions in multiple photometric bands ranging from the ultraviolet to infrared and even (sub-)mm wavelengths yields the spectral energy distribution (SED) and total luminosity of dense cores, young stellar objects and proto-planetary disks. Fitting the SED to a model allows to constrain both stellar and disk properties (see Fig. 7.2 for an example). Spectroscopy in the mid- and far-infrared with space observatories such as Herschel Space Observatory and Spitzer Space Observatory and ground-based (sub-)millimeter interferometer Atacama Large Millimeter Array (ALMA) provide detailed line profiles and velocity maps.

The shape of this energy distribution is closely related to the physical structure of the various objects. Embedded cores, for example, show a single-peaked energy distribution that can be approximated by a single temperature black-body (Fig. 7.3). This is due to the optically thick nature of these objects where all received radiation has been re-processed by the gas and dust of the envelope. A star+disk system in later evolutionary phases displays a clear double peak structure, with the stellar continuum peaking at optical wavelength and the cooler proto-planetary disk at infrared wavelengths. The stellar emission is still partly re-absorbed and scattered by the surrounding circumstellar material (disk and possibly envelope), causing
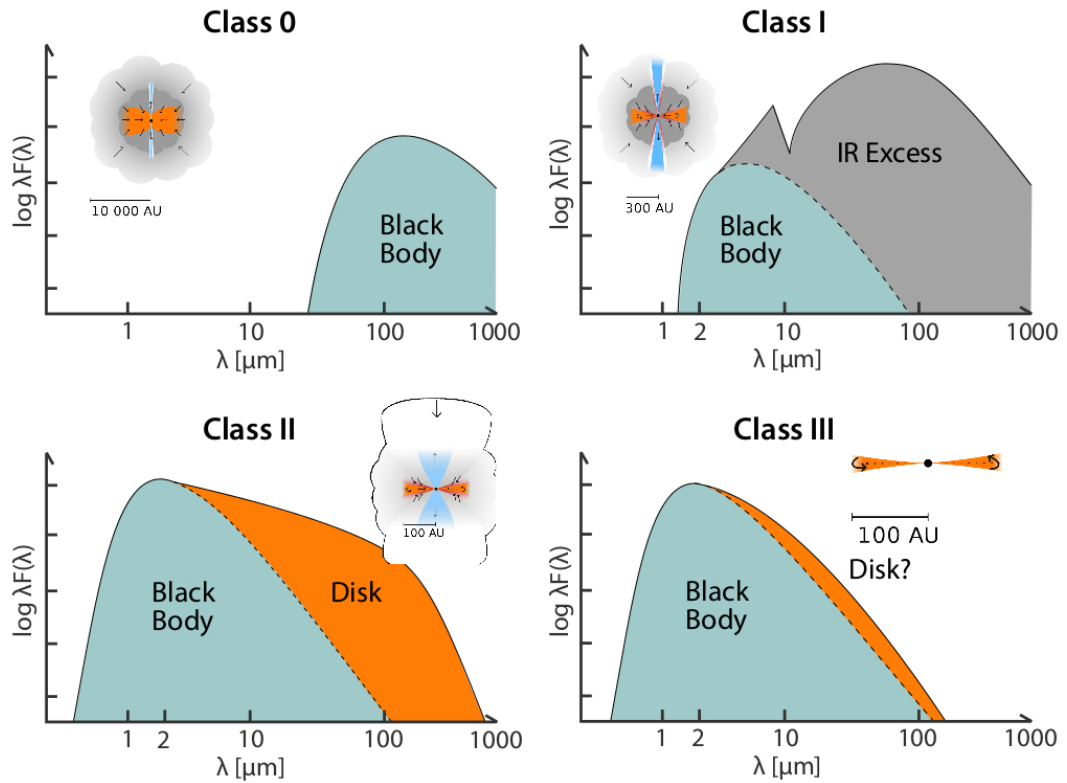
Figure 7.3: *The empirical classification scheme for the appearance of spectral energy distributions (SEDs) of subsequent phases in the star forming process. Class 0 sources have distributions whose widths are similar to single temperature blackbody functions. Class I and Class II sources display infrared excess which produces SEDs which are significantly broader than a single blackbody function. In Class I sources the 10 μm silicate feature is usually in absorption; in Class II sources it may be hard to discern (as shown in the figure) or in emission. Class III sources have only a modest infrared excess and show SEDs that are only slightly broader than a single blackbody. From: Magnus Persson.*

veiling of the star in the UV. The energy removed in this way is reprocessed by the gas and dust in the envelope and re-emitted in the IR.

## 7.2.1 Lada classification

A commonly used observational classification of young objects has been devised by Lada (1999) and involves use of the SED longward of 2 μm. To characterize this SED the slope of the infrared spectra $\alpha_{\mathrm{IR}}$ is defined as

$$\alpha_{\mathrm{R}} = \frac{d \log\left(\lambda \, \mathcal{F}_\lambda\right)}{d \log \lambda} = -\frac{d \log\left(\nu \, \mathcal{F}_\nu\right)}{d \log \nu}, \tag{7.9}$$

in the wavelength range between 2.2 μm and several tens of microns.

**Class I:** So called *Class I* objects have a positive value of the spectral index ($\alpha_{IR} > 0$). The flux density $\mathcal{F}_\nu$ peaks at 60–100 $\mu$m. At 2.2 $\mu$m the flux is increasing toward longer wavelengths, and the spectrum is broader than a black body with little radiation in the near-infrared at $\lambda < 2\,\mu$m.

The physical picture is that a star-disk system has formed. Accretion luminosity heats up the inner disk regions that starts to emit in the near-IR, at wavelengths of roughly a micron. The dust in the (spherical) outer envelope still absorbs a significant fraction of this light, re-emitting it at longer wavelengths. The SED of a Class I source therefore still peaks in the far-IR, but has a significant contribution in the near-IR. This type of object is also known as an 'embedded' IR source, or as a candidate proto-star.

**Class II:** An object of *Class II* has a spectral index $-1.6 < \alpha_{IR} < 0$. The spectrum is also broader than a single black body, with a significant IR excess and also a UV excess. The extra UV continuum radiation causes an apparent 'filling in' or *veiling* of the absorption lines in the blue region of the spectrum.

The outer envelope of Class II objects has become (optically) thin and the star and disk are directly observable. Such objects are interpreted to be pre-main-sequence stars with disks. Class II objects are often T Tauri stars (see below). The spectral energy distribution of Class II sources can be modeled as the sum of that from the now optically visible star together with emission in the infrared and millimeter from a surrounding circum-stellar disk. Ultraviolet emission above the value predicted for a bare stellar photosphere is also observed in Class II sources and is attributed to accretion hotspots on the stellar surface as gas in the disk flows on to the star.

**Class III:** A *Class III* object has $\alpha_{IR} < -1.6$. Its spectrum is close to that of a black body of a single temperature, and there is little or no evidence for excess IR radiation or dust.

These objects are either pre-main-sequence stars beyond the T Tauri phase, i.e. without indications of significant disks, or young main-sequence stars. Some also display evidence for surrounding *debris-disks* – gas poor disks whose emission is attributed to short-lived dust that is continually regenerated by erosive collisions of larger solid bodies in orbit around the star.

Though not in the original classification of Lada, also *Class 0* objects have been identified. They are thought to correspond to proto-stars at an even earlier evolutionary phase than that represented by Class I. Some of the radiation is in the sub-millimeter, with hardly any radiation shortward of 10 $\mu$m (hence the reason why these sources do not fit in the original Lada classification).

**Class 0:** Their spectral energy distributions look like black bodies with temperatures of 15–30 K and a peak near 100 $\mu$m, creating an SED that is very red in color.

One of the commonly used criteria for Class 0 is that the mass in the in-falling region of such objects is greater than the mass in the central hydrostatic region, while the reverse
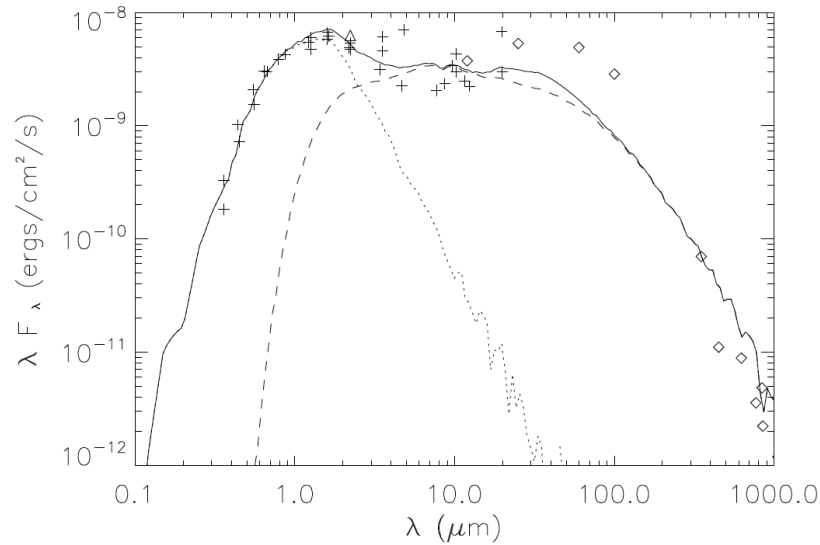
Figure 7.4: *The spectral energy distribution of the young star T Tauri and model curves. T Tau has a spectral slope of its infrared spectrum in the 2.2-tens of μm range that is almost 'flat'. This implies that it is actually intermediate between a Class I and Class II spectrum and is sometimes assigned the class classification* Flat. *Things are usually more complex than anticipated: T Tau is actually a triple system. The binary component, T Tau S, consisting of T Tau Sa and T Tau Sb (separation ∼ 0".1), is currently ∼ 0".7 south of T Tau N. All three stars are actively accreting and believed to host disks (Duchêne et al. 2005). Another possibility is that T Tau S is obscured by the disk around T Tau N. Plus-signs indicate resolved data of T Tau N, diamonds indicate where T Tau S and T Tau N are unresolved and the flux of T Tau S dominates. Solid line: total flux; dotted line: stellar flux; dashed line: disk flux. From: Gustafsson et al., 2008, A&A 488, 235.*

is true in Class I. Bok globules that contain an optically thick core are Class 0 objects. The core is completely hidden by the (optically) thick, cold dusty envelope around it.

Another extension of this classification is that of a *Flat* spectral class, for stars that have a spectral slope $\alpha_{IR} \sim 0$. T Tau is an example of this class, the SED of which is shown in Fig. 7.4. We can now summarize the classification as follows (see also Green et al., 1994, ApJ 434, 614):

| | |
|---|---|
| 0 | $\alpha_{IR}$ undefined |
| I | $\alpha_{IR} > +0.3$ |
| Flat | $-0.3 \leq \alpha_{IR} < +0.3$ |
| II | $-1.6 \leq \alpha_{IR} < -0.3$ |
| III | $\alpha_{IR} < -1.6$ |

Note that the quantity $\alpha_{IR}$ is not useful in connection to Class 0. In fact, the boundaries between the classes of objects are somewhat fuzzy.
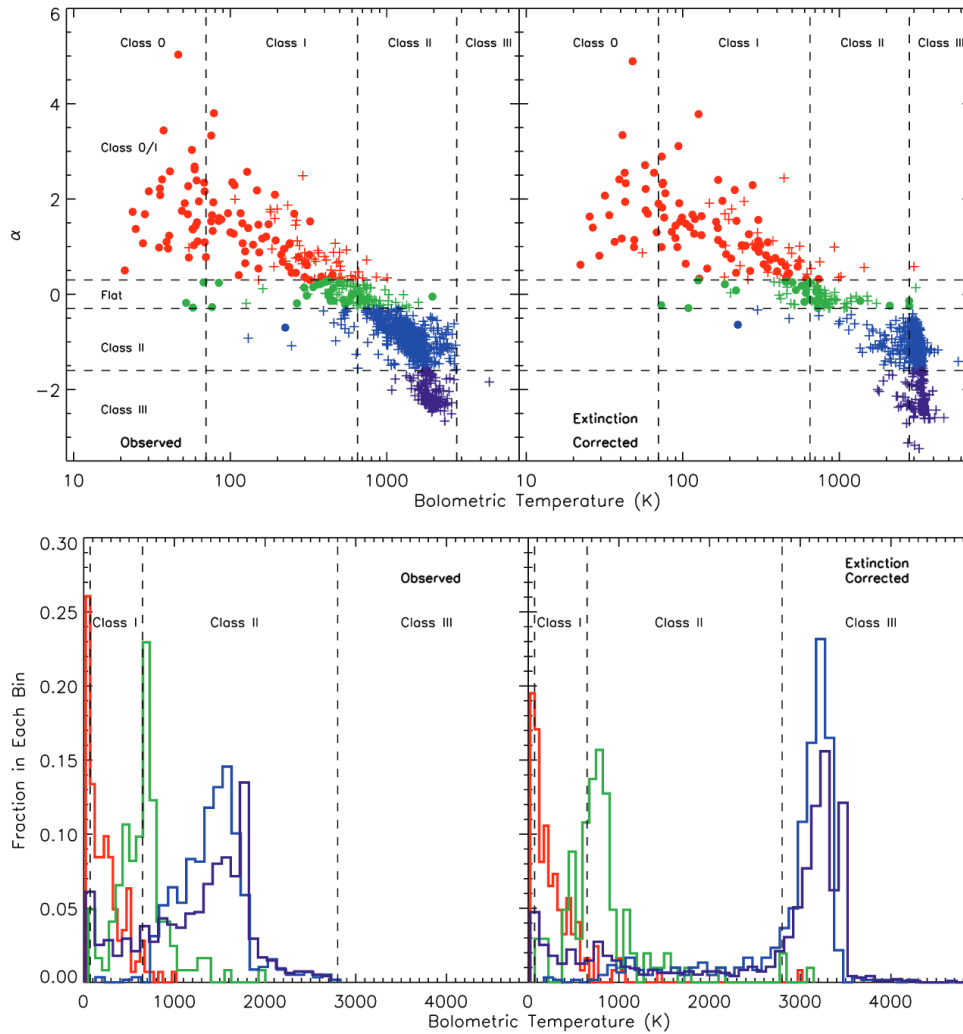
Figure 7.5: *Top: The value of $\alpha_{\rm IR}$ is plotted versus $T_{\rm bol}$ for sources with well-determined values in five nearby star-forming regions: Cham. II, Lupus, Perseus, Serpens, and Ophiuchus. The color code is based on the Lada class of each source, as defined by $\alpha$, with Class I plotted as red, flat as green, Class II as blue, and Class III as purple. Filled circles indicate sources associated with envelopes as traced by millimeter continuum emission, while plus signs indicate sources with no such associations. The right panel shows $\alpha'$ and $T'_{\rm bol}$ calculated after corrections for extinction were applied. The vertical dashed lines show the boundaries between classes, as defined by Chen et al. (1995). Bottom: Distribution of $T_{\rm bol}$ for sources in each of the Lada classes, in linear $T_{\rm bol}$ bins. The color code is the same as in the top panel. The right panel shows the same distributions after corrections for extinction were applied to the observations. Figures and captions taken from Neal J. Evans et al. (2009, ApJS 181, 321).*

With the large amounts of SEDs secured with the *Spitzer Space Telescope*, the onset of the IR excess $\nu_{\rm onset}$ (or $\lambda_{\rm onset}$) is used as a second parameter to characterize the energy distribution. It is defined as the wavelength where the measured fluxes start to be larger than the pure

photospheric value. This parameter proved extremely useful in sub-classifying the rather in-homogenous Class II sources, i.e. star+disk systems (see Fig. 7.5).

If one takes a census of the relative numbers of objects in Class 0, Class I, the Flat class, and Class II in a given star formation region, one can make an estimate of the relative amounts of time spent in these four phases. Early studies of the Taurus region and the $\rho$ Ophiuchi cloud indicate that the embedded proto-stellar phase (basically Class I) lasts 0.2–0.3 Myr (Kenyon et al. 1990, AJ 99, 860; Wilking et al. 1989, ApJ 340, 823). A more extensive survey of star-forming regions with the Spitzer Space Telescope gives an average value of about 0.1, 0.44, and 0.35 Myr, respectively, for Class 0, Class I, and the Flat class (Evans et al. 2009, ApJS 181, 321; see also Fig. 7.5). The embedded phase is then followed by a (classical) T Tauri phase (Class II) that lasts, on the average, 1–3 Myr. However, these numbers are un-certain; sources of uncertainty include incompleteness, contamination by background objects, extinction corrections, and the assumption that star formation is occurring at a constant rate, allowing a continuous flow through the various classes.

## 7.3 Evolution in the HR diagram

Figure 7.6 shows the evolutionary tracks that pre-main-sequence stars of various masses take before they reach the hydrogen burning stage. The start of H-burning is referred to as the *zero-age main sequence* or ZAMS. It is the left-most solid line – i.e., where stars are the most compact – connecting the end points of the tracks for different masses. Above the figure a table is given with timescales for various parts of the formation process. A 0.6 $M_\odot$ star takes 85 Myr to reach the ZAMS, solar mass stars 32 Myr, and $> 4$ solar mass stars less than 1 Myr.

### Birthline

Fig. 7.6 also contains *isochrones*, i.e. lines of constant age since the star passed the *birthline*, defining $t = 0$ yr. In the definition that we will adopt, the top solid line connecting the starting points of the tracks is the birthline (note that the isochrone of $10^5$ yr virtually coincides with the birthline for the lowest masses). However, be ware that there is no unique definition, which is cause for substantial confusion. Even worse, theoreticians and observers apply different concepts. Let us first focus on alternative theoretical definitions (from ours). Most theorists define the birthline in terms of the behavior of the mass accretion rate. Maeder (2009) defines the birthline to be the track that a proto-star describes in the HR diagram while it is accreting at a 'substantial rate'. Once this accretion stops, or at least strongly diminishes, the star 'leaves' the birthline. What is a 'substantial rate' of mass accretion? An expression for this may be derived requiring the Kelvin-Helmholtz contraction time to be much larger than the accretion time, i.e. $\tau_{\text{KH}} \gg \tau_{\text{acc}}$ (see Sect. 6.5). This yields

$$\dot{M} \gg \frac{RL}{GM} = 6.4 \times 10^{-7} \left( \frac{R}{2R_\odot} \right) \left( \frac{L}{10L_\odot} \right) \left( \frac{M}{M_\odot} \right)^{-1}. \qquad (7.10)$$

TABLE 1

CHARACTERISTIC TIMES OF PRE–MAIN-SEQUENCE EVOLUTION

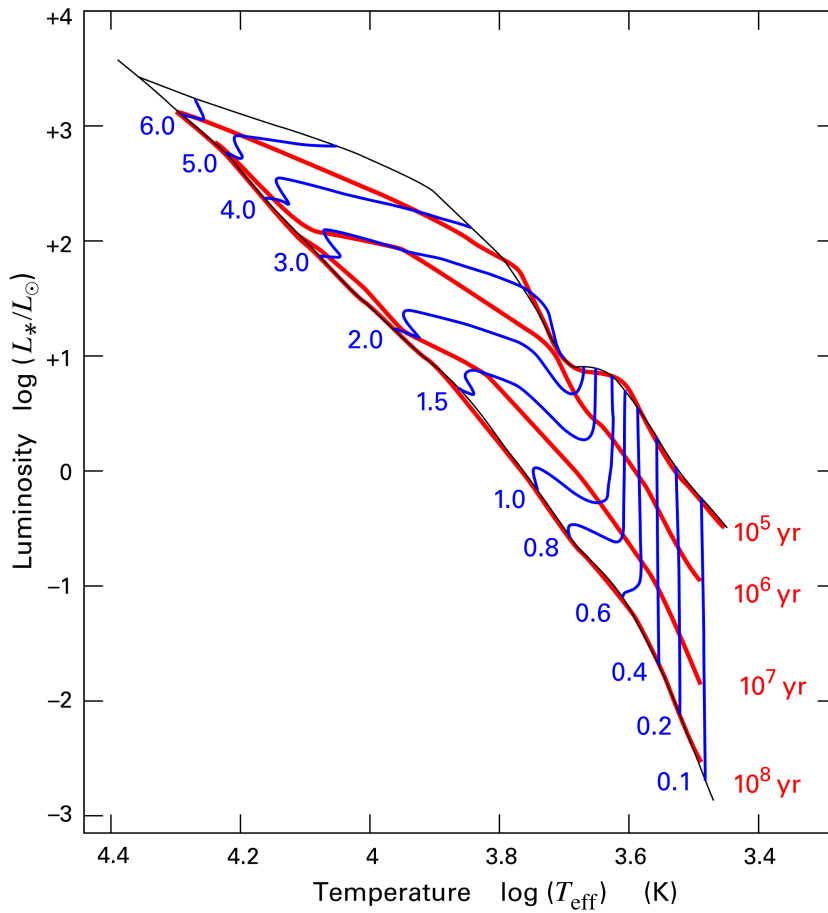| Mass ($M_\odot$) | $\Delta t_D$ (yr) | $t_{rad}$ (yr) | $t_{ZAMS}$ (yr) |
|---|---|---|---|
| 0.1...... | $1.5 \times 10^6$ | ... | $3.7 \times 10^8$ |
| 0.2...... | $7.0 \times 10^5$ | ... | $2.4 \times 10^8$ |
| 0.4...... | $3.0 \times 10^5$ | $1.1 \times 10^7$ | $1.1 \times 10^8$ |
| 0.6...... | $2.0 \times 10^5$ | $5.5 \times 10^6$ | $8.5 \times 10^7$ |
| 0.8...... | $1.5 \times 10^4$ | $2.5 \times 10^6$ | $5.2 \times 10^7$ |
| 1.0...... | ... | $1.4 \times 10^6$ | $3.2 \times 10^7$ |
| 2.0...... | ... | ... | $8.5 \times 10^6$ |
| 3.0...... | ... | ... | $2.0 \times 10^6$ |
| 4.0...... | ... | ... | $8.2 \times 10^5$ |
| 5.0...... | ... | ... | $2.3 \times 10^5$ |
| 6.0...... | ... | ... | $4.0 \times 10^4$ |



Figure 7.6: *Pre-main sequence evolutionary tracks of low-mass stars in the HR-diagram. Each track is labeled by the corresponding stellar mass, in solar masses. The tracks are shown from the moment that the object passes the birthline (defining $t = 0$), which is the top solid black line. Selected isochrones are shown by the red lines. The tracks end when the objects settles on the zero age main sequence or ZAMS. The table lists timescales: $\Delta t_D$ is the duration of deuterium burning; $t_{rad}$ the time at which a radiatively stable core first appears in fully convective models, and $t_{ZAMS}$ the total time required for each star to settle onto the ZAMS. From: Palla & Stahler 1999, ApJ 525, 772*
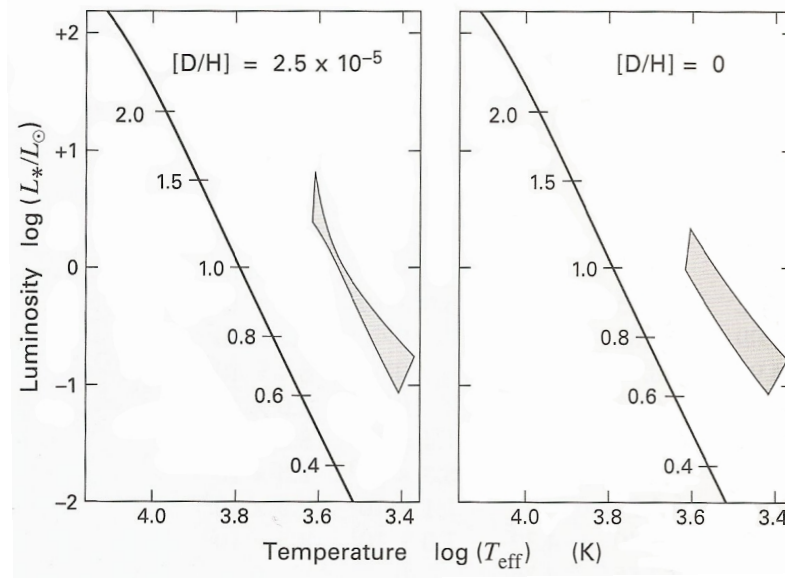
Figure 7.7: *Effects on the theoretical birthline of changes in the mass accetion rate $\dot{M}$. The upper curve in each case is the birthline constructed with $\dot{M} = 1 \times 10^{-5} \, M_\odot \mathrm{yr}^{-1}$, while the lower curve corresponds to $2 \times 10^{-6} \, M_\odot \mathrm{yr}^{-1}$. In the left panel, the calculations assume the full interstellar deuterium abundance, while this abundance has been artificially set to zero in the right panel. From: Stahler & Palla, The Formation of Stars.*

Note that this mass loss is about the mass-accretion rate Eq. 6.8 in the Shu inside-out collapse model. This thus implies that the star leaves its birthline essentially when the mass accretion is quenched because all of the cloud material is consumed. Realize that during the main accretion phase the protostar, as seen by an outside observer, can not be placed in the HRD because the radiation leaving its surface will be absorbed by the dust envelope, degrading the surface photons to infrared radiation (see Fig. 6.7). Once the cloud material is used up, the dust envelope will be accreted or dispersed, hence the star becomes optically visible. Therefore, in this definition, the end of the birthline is the moment the star becomes a pre-main sequence star. Our definition is different in the sense that the birthline does not refer to a track of a single object, but to the locus of points connecting stars that have reached the end of their main-accretion phase. The birthline gives the HRD position of stars of a range of masses at the end of a prescribed mass-accretion history. The tracks shown in Fig. 7.6 have been computed for a constant accretion rate $\dot{M}_{\mathrm{acc}} = 10^{-5} \, M_\odot \mathrm{yr}^{-1}$. Once the amount of mass that is labeled at the ZAMS position is accumulated, accretion stops (the cloud being consumed) and the star is revealed to the eye, i.e. becomes optically visible.

Predictions of the location of the birthline depend on the results of both proto-star and pre-main-sequence theory. Specifically, the precise value of $\dot{M}$, its temporal behavior, or even the geometric pattern of infall, influence the birthline location. Interestingly, they do not strongly influence the curve's location. How does the birthline respond to a shift in the infall rate? For a

lower $\dot{M}$ the radius, at a given mass, diminishes, which shifts the birthline down and to the left in Fig. 7.6 (see Fig. 7.7). Slower buildup of its mass gives the proto-star more time to radiate internal energy and therefore shrink. Nevertheless, the radius change is severely restricted at sub-solar masses. The reason, we recall, is the thermostatic effect of core deuterium fusion. Although present in small quantities, this fuel effectively constrains the radius by altering the nuclear energy output during the main accretion phase. In calculations where deuterium fusion is artificially suppressed, the mass-radius curve (shown in Fig. 6.9 for an accretion rate of $10^{-5}$ $M_\odot \mathrm{yr}^{-1}$) undergoes a larger shift when $\dot{M}$ is altered.

Figure 7.7 shows the effect on the birthline of altering the infall rate. As expected, inclusion of deuterium assures a relatively narrow spread in the predicted birthline. In contrast, the maximum luminosities for pre-main-sequence stars would have a larger dispersion in the hypothetical situation of zero deuterium abundance. This reasoning implies, incidentally, that any rise in the estimated deuterium abundance relative to hydrogen in the interstellar medium, a matter of ongoing investigation and here taken to be [D/H] = $2.5 \times 10^{-5}$, would tighten the theoretical boundary even more. The other issues we raised – the temporal behavior of $\dot{M}(t)$ and the spatial pattern of infall – have not yet been addressed through detailed calculations. It is anticipated, however, that these effects are modest or even minor.

Observers define the birthline by identifying the lower luminosity limit of a region in the HRD above which no pre-main sequence stars are observed. Figure 7.10 shows the location of PMS and indeed the lower limit of this 'forbidden zone' quite nicely agrees with the theoretical birthline limit that we adopt – which, thinking about it, is actually about what we should have expected given the theoretical definitions above.

## 7.4  PMS classification

Fig. 7.8 provides an overview of the various evolutionary phases of star and disk formation in the context of object classification, physical and observational properties. In the following, we discuss briefly four classes of pre-main sequence objects, three low-mass object types, the classical T Tauri, weak-line T Tauri, and FU Orionis stars, and the intermediate-mass object type, the Herbig AeBe stars.

### Classical T Tauri stars

*Classical T Tauri stars* (CTTS) are named after the prototype T Tau in the Taurus star forming region. Characteristics are strong Balmer line emission (H$\alpha$), and an association with reflection nebulae and molecular clouds. These stars are low-mass pre-main sequence stars with spectral types between F and M and effective temperatures ranging from 3000 K to 7000 K. They are Class II sources according to their SED and the IR excess peaks between 1 and 10 $\mu$m. The median disk mass is 0.01 M$_\odot$ with a huge spread around it (ranging from
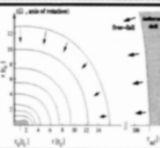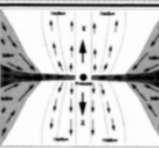
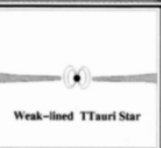| | Infalling protostar | Accreting protostar | Contracting PMS star | | MS star |
|---|---|---|---|---|---|
| **YSO properties** | | | Classical TTauri Star | Weak–lined TTauri Star | |
| **Phase** | adiabatic (A,B,C) | accretion (D) deuterium burning onset of convection | convective radiative onset of nuclear burning | | convective radiative full nuclear burning |
| **Matter flows** | mostly infall disk & outflows form | some infall mostly accretion outflows, jets | low accretion | ? | — |
| **Envelope/ disk size** | < 10000 AU | < 1000 AU | < 400 AU | ~ 100 AU | — |
| **Infall/ accretion rate** | $10^{-4}$ | $10^{-5}$ | $10^{-6} -- 10^{-7}$ | ? | — |
| **Age** | $10^4 - 10^5$ yr | $10^5$ yr | $10^6 -- 10^7$ yr | $10^6 -- 10^7$ yr | — |
| **Emission bands** (except IR) | thermal radio X–ray? | radio X–ray | radio optical strong X–ray | non–therm. radio optical strong X–ray | non–therm. radio optical X–ray |
| **Classes** | Class 0 | Class I | Class II | Class III | ZAMS |

Figure 7.8: *The IR classification in the context with evolutionary phases and matter flow parameters for stars up to $\sim 2\ M_\odot$. The infall rate is in $M_\odot \mathrm{yr}^{-1}$. From: Schulz 2005.*

$0.0003 - 1\ M_\odot$). Typical disk radii are around 100 AU (see Fig 7.9 for a sketch of the star-disk system). Accretion shocks and stellar activity (coronal heating) produce strong X-rays with luminosities up to $L_X \simeq 10^{31}$ erg s$^{-1}$ or $0.003\ L_\odot$.

## Weak-lined T Tauri stars

Young low-mass stars that are associated with star forming regions, but show weak H$\alpha$ emission are generally classified as *weak-lined T Tauri stars* (WTTS). A quantitative criterium is for the equivalent width of the H$\alpha$ line to be smaller than 10 Å. In terms of their SED classification, they often end up as Class III sources because they lack a near-IR excess (due to the absence of an inner disk). It is often assumed that they are the older counterparts of the classical T Tauri stars. Even though there may still be a tenuous disk present in these systems, there is no clear signature of accretion or optically thick dust emission. Since most of the X-ray emission from low-mass stars originates from stellar activity and not from accretion, WTTS are strong X-ray sources just like the CTTS.
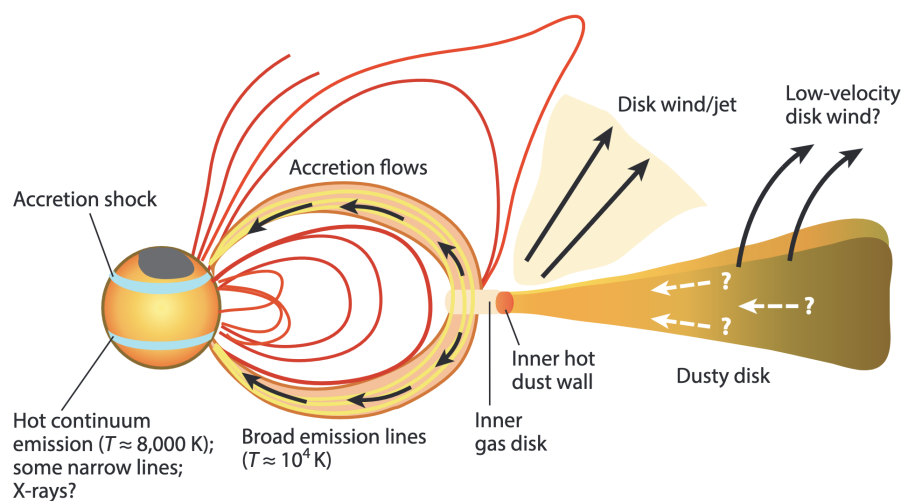
**Figure 7.9:** *Schematic picture of accretion onto a young star from a circum-stellar disk. The strong stellar magnetic field produces large starspots and disrupts or even truncates the disk at a few stellar radii. Magnetic field lines connected with the disk channel material at near free-fall velocities (causing broad emission lines) that then shocks at the stellar surface (producing a hot continuum). Other magnetic field lines unconnected to the disk produce coronal X-ray emission and/or a stellar wind. Some field lines may become twisted by differential rotation between disk and star or in the inner (Keplerian) disk, causing field lines to bulge out or even eject matter. The inner 1 AU of the disk produces a bipolar flow or jet, driven by accretion energy. An MHD wind may also be needed to drive disk accretion at larger radii. The disk itself emits at infrared, sub-mm and mm wavelengths, and extends up to $\sim 100$ AU. So, the figure elements are not to scale. From: Hartmann et al. (2016).*

## FU Orionis and EX Orionis stars

*FU Orionis or FUor type stars* have spectral types ranging from F to G with effective temperatures of $\sim 6000$–7000 K. They show periodic optical outbursts of several magnitudes with a rising timescale of $\sim 1$ yr and a decline timescale of 50–100 yr (see Fig. 7.11). It seems that the outbursts are connected to episodes of increased mass accretion rates from the disk (up to $10^{-4}$ M$_\odot$/yr), possibly related to dynamical instabilities. A more extensive description of the nature of these objects, the underlying accretion physics and observational properties can be found in Hartmann et al. (2016).

Are FU Orionis outbursts rather unique events affecting only a limited number of pre-main-sequence stars? The fact that in less than a century about a dozen such sources were seen, suggests that it must be a widespread phenomenon affecting probably all PMS stars several times during their evolution. Indeed, if one compares the amount of known FU Orionis objects with that of all PMS stars within the same volume, and considers the expected duration of the PMS evolution, the conclusion is quite clear that such outbursts must be a recurrent phenomenon in virtually all PMS stars.
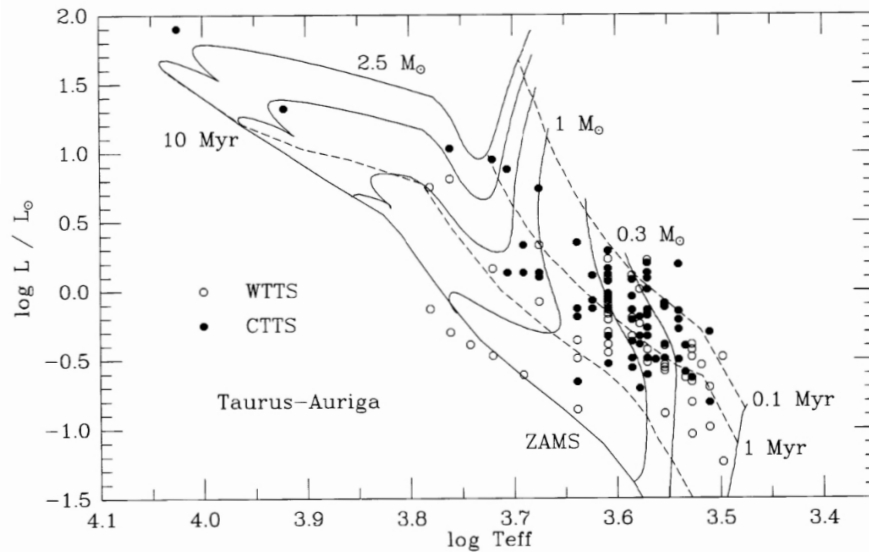
Figure 7.10: *HR diagram positions of young stars in the Taurus-Auriga molecular cloud complex. For comparison, theoretical evolutionary tracks for pre-main-sequence stars of masses 2.5, 2.0, 1.5, 1.0, 0.5, 0.3, and 0.1 $M_\odot$ are shown. The dashed lines are isochrones for ages of $10^5$, $10^6$, and $10^7$ yr with the hydrogen-fusion zero-age main sequence (ZAMS) shown as the lowest line running from upper left to lower right. The open circles represent weak-emission T Tauri stars (WTTS) and the filled circles the classical T Tauri stars (CTTS). Note that the WTTS appear to be closer to the ZAMS than the CTTS, consistent with the idea that WTTS represent a more evolved state of PMS evolution than do the CTTS. From: Hartmann 1998.*

It is likely that the spectacular FU Orionis outbursts constitute the tail in the distribution of more frequent discontinuities in the mass accretion rate, and that more frequent smaller-amplitude events also occur. In fact, the irregular low-amplitude variations on timescales of days of virtually all T Tauri stars may represent the other limit of the distribution, being caused by stochastic variations of the accretion rate. In between, intermediate-amplitude outburst-like events with a characteristic repeat time of less than a year termed *EX Orionis or EXor* may fit. These have been named after the recurrent rather sizeable flares which objects such as EX Lupi exhibit (Herbig 1989).

## Herbig AeBe stars

*HAeBe stars* have intermediate masses between 2 and 8 $M_\odot$. They are the more massive counterparts of the T Tauri stars. The current definition can be summarized as stars with spectral types between B0 and early F, luminosity classes II to V (do not get confused with the Lada classification, here is meant the Morgen-Keenan classification), having emission lines (e.g. strong H$\alpha$) in their spectra and showing an infrared excess due to the presence of circumstellar material. The pre-main-sequence evolutionary timescales of Herbig AeBe stars
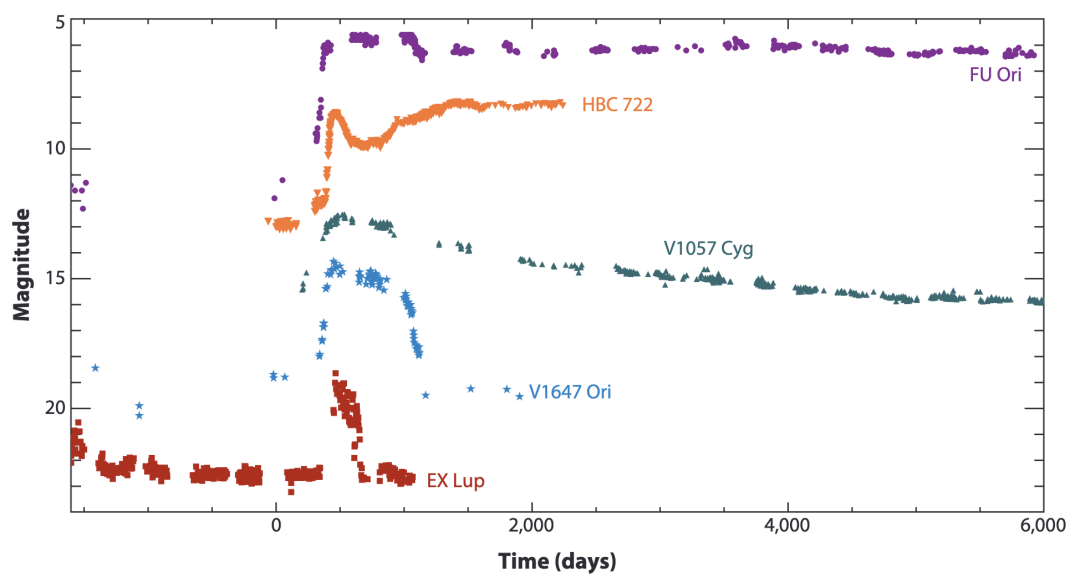
Figure 7.11: *Lightcurves of FUor outbursts of FU Ori and HBC 722; EXor outbursts of EX Lup and V1647 Ori; and an outburst that is intermediate to these two outburst types in V1057 Cyg. From: Hartmann et al. (2016); in turn adapted from Kóspál et al. (2011).*

(also often called HAEBE stars) is much shorter than that of T Tauri stars. In the early phases, these stars are only relatively briefly fully convective and, predominantly, undergo a phase of thermal relaxation. Hence, they mostly lack the Hayashi part of the evolutionary track. Herbig AeBe stars initially move upward in the HR diagram due to their expansion (see also right side panel of Fig. 6.9).

## Exercise 7.1

Compute the slope $\alpha_R$ for the Rayleigh-Jeans limit of a black body curve.

## Exercise 7.2

HAeBe stars show excess of continuum emission compared with main-sequence stars with similar spectral types, in particular at wavelength short-ward of the Balmer jump (at 364.6 nm or 3646 Å). This Balmer-continuum excess is ascribed to a localized heating of the surface due to magnetospheric accretion of gas from its circumstellar disk. Figure 1 shows the effect. The spectral energy distribution of the star is modelled in terms of (1) the fraction $f$ of the surface where material falling in from the inner radius $R_{in} = 2.5\,R_\star$ of the circum-stellar disk causes accretion shocks, heating the local surface; (2) the temperature $T_{col}$ (see below), and (3) the mass-accretion rate $\dot{M}_{acc}$. We assume that the luminosity in the columns in which accretion takes place is given by

$$L^{col} = f 4\pi R_\star^2\, \mathcal{F}^{col} = f 4\pi R_\star^2 \left(\mathcal{F}^{acc} + \mathcal{F}^{phot}\right), \qquad (7.11)$$

where $R_\star$ is the stellar radius and $\mathcal{F}^{col} = \sigma T_{col}^4$ is the total column flux, with $T_{col} = 12\,470$ K. $\mathcal{F}^{acc}$ is the accretion flux, i.e. the flux generated by the conversion of potential energy of accreting gas into radiation. The total photospheric flux $\mathcal{F}^{phot} = \sigma T_\star^4$, where $T_\star = 9\,000$ K. The stellar mass is $M_\star = 2.5\,M_\odot$; the stellar radius $R_\star = 2.6\,R_\odot$.

a) Introduce the accretion luminosity $L_{acc}$ and give an expression for the actual luminosity that is generated by the material that is falling onto the star, explicitly introducing $L_{acc}$. Call this luminosity $L_{acc}^{real}$.

b) How are the total accretion flux $\mathcal{F}^{acc}$ and the accretion luminosity $L_{acc}$ related?

c) Give an expression for the mass accretion rate $\dot{M}_{acc}$ in terms of relevant stellar and disk parameters, and the parameters $f$ and $T_{col}$.

d) Calculate the mass accretion rate in $M_\odot\mathrm{yr}^{-1}$ for the star described in Fig. 7.12.

e) If only U and B photometry are available (that should match the filled circles in Fig. 7.12), is it possible to constrain both $T_{col}$ and $f$ in the fitting procedure (or only one of the two)? Explain your answer.

## Exercise 7.3

The Mendigutia et al. 2011 study discussed in the previous exercise sets out to constrain the mass accretion rates from violet (Balmer-continuum) excess for a total of 38 HAeBe stars. For these stars the authors also measure the equivalent widths $W_{eq}$ of the H$\alpha$,
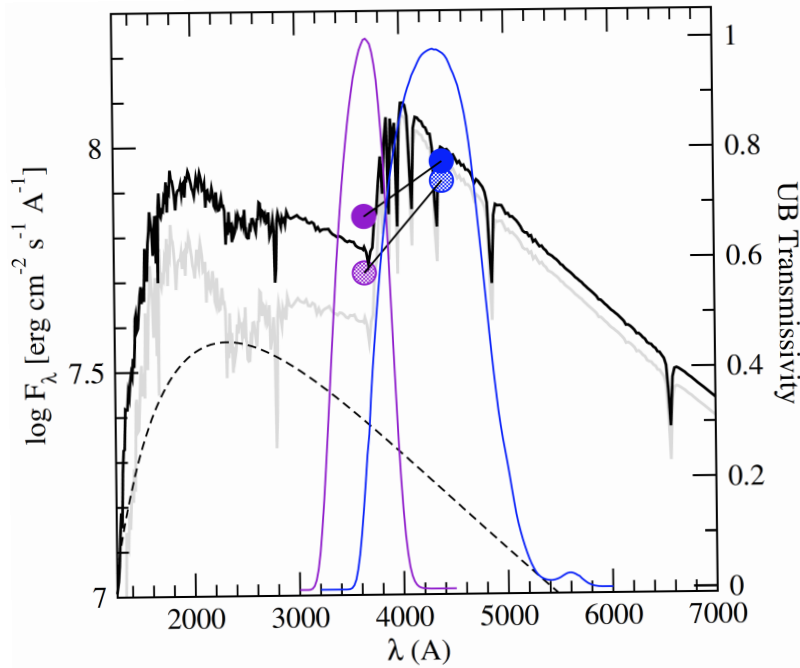
Figure 7.12: *Fit to the U and B photometry of a Herbig Ae star using a two component model. The first component (in grey) is the photospheric flux, for which a Kurucz spectral energy distribution is used with $T_\star = 9\,000\,K$ and $g_\star = 10^4\,cm\,s^{-1}$ (or $M_\star = 2.5\,M_\odot$ and $R_\star = 2.6\,R_\odot$). The second component is a Planck spectrum of temperature $T_{\rm col} = 12\,470\,K$, shown with a dashed black line. The two are added using a relative weighing factor $f = 0.084$ such that the the total flux (shown with the solid black line) is given by $F_\lambda = f F_\lambda^{\rm col} + (1 - f)F_\lambda^{\rm phot}$. Here, $F_\lambda^{\rm col}$ is the flux generated in the columns in which material from the inner disk rim lands on the surface. The total surface covered by these columns is a fraction $f$ of the stellar surface. All fluxes are computed at the stellar surface. The transmission curves of the U and B filters are plotted in violet and blue, respectively. The convolution of the fluxes with the filter responses provide the synthetic photometric points used to obtain the modelled photospheric and total colors (from the shaded and filled circles, respectively). From: Mendigutia et al. 2011, A&A 535, A99.*

[O I]$\lambda$6300, and Br$\gamma$ emission lines they observe in their data. They correct these equivalent widths for the photospheric contribution to the line strengths using atmosphere models. See e.g. the model H$\alpha$ line at $\lambda$6563 Å in Fig. 7.12. Such a correction isn't necessary in the case of [O I] as this line has no photospheric component. Then, they compute the line integrated flux by multiplying $W_{\rm eq}$ by the observed (interstellar extinction corrected) continuum flux just next to each of these lines. Having done this, they can obtain the line luminosity from

$$L(\text{line}) = 4\pi d^2\,|W_{\rm eq,obs}(\text{line}) - W_{\rm eq,int}(\text{line})|\,\mathcal{F}_\lambda^{\rm cont,obs}, \tag{7.12}$$

where $d$ is the distance.

a) They hypothesize that a correlation should exist between the line luminosity $L(\text{line})$ and the accretion luminosity $L_{\rm acc}$. Explain why this is a reasonable expectation.

They check their hypothesis by plotting the two properties against each other and indeed find that they correlate. For the H$\alpha$ line the empirical relation is

$$\log\left(\frac{L_{\mathrm{acc}}}{L_\odot}\right) = 2.28(\pm0.25) + 1.09(\pm0.16) \times \log\left(\frac{L(\mathrm{H}\alpha)}{L_\odot}\right). \qquad (7.13)$$

For weak-lined T Tauri stars similar relations have been derived by Muzerolle et al. 1998, Dahm 2008, and Herczeg & Hillenbrand 2008.

# 8

# The formation of massive stars

Stars with masses larger than $\sim$8–9 $M_\odot$ reach the main sequence before they become optically visible and before they have ended the main accretion phase. Note indeed in Fig. 7.6 that the birthline merges into the ZAMS at these masses. At all times in the formation process such objects are therefore severely obscured and hard to study observationally.

It is an interesting coincidence (but not more than that) that the mass limit of $\sim$8–9 $M_\odot$ coincides with a major change in the post-main sequence evolution of stars; such objects do not develop degenerate cores and therefore end their lives in core-collapse supernovae. Stars above 8–9 $M_\odot$, i.e. stars that start out their lives as B2 stars or earlier, are referred to as *massive stars*. The most reliable mass measurements of massive stars are from those in binary systems with dynamically measured masses. In the Galaxy the most massive binaries currently known are NGC3603-A1 (116 $\pm$ 31 and 80 $\pm$ 16 $M_\odot$; Schnurr et al. 2008), WR20a (83 $\pm$ 5 and 82 $\pm$ 5 $M_\odot$; Rauw et al. 2005; Bonanos et al. 2004), and WR21a (104 $\pm$ 10 and 58 $\pm$ 4 $M_\odot$; Niemela et al. 2008; Tramper et al. 2016). From analysis of the intrinsically brightest stars in massive clusters it appears that even more massive stars exist. The current record holders – discovered by de Koter et al. (1997, 1998) – are located in the central cluster of the 30 Doradus region, Radcliffe 136a (R136a; see Table 2.3 and Fig. 2.8). For objects R136a1, a2, and a3, Brands et al. (2022) estimated masses of $222^{+28}_{-29}$, $186^{+17}_{-15}$, and $179^{+16}_{-11}$ $M_\odot$, respectively (see Fig. 8.1). Probably these stars are about 1–2 Myrs old and have lost part of their initial mass in an intense outflow or stellar wind. Correcting for this, the initial masses of the stars were probably $273^{+25}_{-36}$, $221^{+16}_{-12}$, and $213^{+12}_{-11}$ $M_\odot$. How could such massive stars have formed?

Two main routes towards the formation of massive stars have been proposed: (1) formation via gas accretion, i.e. the same mechanism by which stars of low mass form, and (2) formation via collisions between lower mass stars, which would represent a very different formation mechanism from the bulk of the stellar population. We discuss both mechanisms.
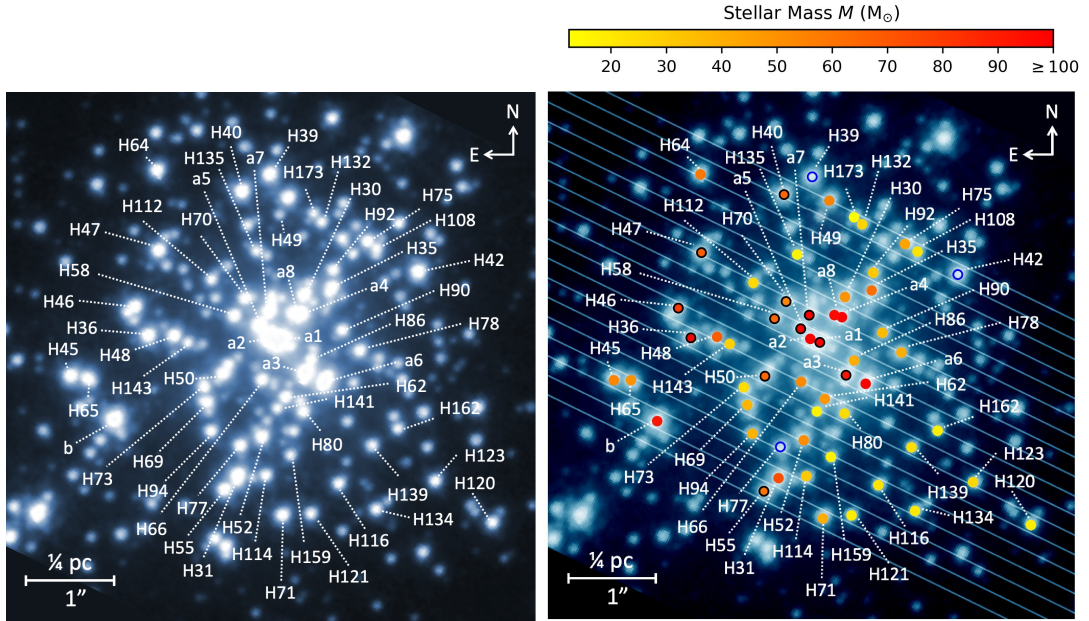
Figure 8.1: *The central region Radcliff 136a of the 30 Doradus region in the Large Magellanic Cloud observed in the V-band (filter F555W) with HST/WFC3.* Left: *Star identifications starting with 'H' from Hunter et al. (1995); those starting with 'a' and 'b' from Weigelt & Baier (1995). The three most massive stars are a1, a2, and a3.* Right: *The colour of the circles indicates the current (evolutionairy) mass of each source. Notice the crowding of the stellar sources, especially in the very core regions. From: Brands et al. 2022.*

## 8.1 The formation of massive stars by accretion

There are several major challenges to forming massive stars by accretion, which we will discuss below: fragmentation, mass-accretion rate, radiation pressure, ionization feedback, and stellar winds.

### Fragmentation

Equation (3.61) provides a convenient expression for the Jeans criterion, i.e. the minimum mass necessary to initiate the collapse of a cloud due to its self-gravity. To aid the discussion, we repeat it here

$$M_{\mathrm{J}} = 7.41 \left( \frac{T}{10\,\mathrm{K}} \right)^{3/2} \left( \frac{2}{\mu} \right)^2 \left( \frac{10^4}{n} \right)^{1/2} M_\odot. \tag{8.1}$$

For a typical dark cloud with $T = 10\,\mathrm{K}$, $\mu = 2.3$, and $n = 10^4\,\mathrm{cm}^{-3}$ (see also Table 2.2) $M_{\mathrm{J}} = 5.6\,M_\odot$. Clearly, a larger cloud with similar local conditions would fragment imme-

diately. What processes, then, can increase the Jeans mass, i.e. reduce the fragmentation? Fragmentation is reduced by two primary effects: radiation feedback and magnetic fields.

Radiation feedback works to reduce fragmentation by heating the gas, raising its pressure and thus its Jeans mass. For instance, if the temperature of the medium would be increased to $200\,\mathrm{K}$, the Jeans mass equation would imply $M_\mathrm{J} \sim 500\,M_\odot$. We address the question of how to heat the medium below, when discussing the mass-accretion rate.

Magnetic fields limit fragmentation in two ways. First, they remove angular momentum. In a collapsing cloud, the densest regions have a shorter free fall time (Eq. 3.79) and collapse fastest. As the gas falls inward it will rotate faster and faster in order to conserve angular momentum. When the collapsing gas is threaded by a magnetic field, the resulting differential rotation between inner collapsing regions and outer ones twists the magnetic field lines. The twist produces a magnetic tension force (see Eqs. 3.34 and 3.35) that transfers angular momentum from the inner to the outer regions, a process known as *magnetic breaking*. For an axi-symmetric flow one can show (e.g. Stahler & Palla 2005) that the time rate of change of the angular momentum of a fluid element at a distance $r$ from the rotation axis due to magnetic forces is given by

$$\frac{\partial J}{\partial t} = \frac{1}{4\pi}\left[B_r\frac{\partial}{\partial r}(rB_\phi) + rB_z\frac{\partial}{\partial z}B_\phi\right],\tag{8.2}$$

where $\mathbf{B} = (B_r, B_\phi, B_z)$ is the magnetic field, and we have used cylindrical coordinates (therefore, beware of the definition of $r$). For the types of magnetic field configuration produced by collapse, the net effect is to transport angular momentum outward.

This process inhibits the formation of rotationally-flattened structures such as accretion disks. This is of consequence from the standpoint of fragmentation, because rotational flattening raises the density of the gas as it approaches the star, and dense, rotationally-flattened structures are vulnerable to the Toomre instability (see Eq. 9.63), in which the self-gravity of a flattened rotating structure overcomes support from thermal pressure and angular momentum, leading to fragmentation and collapse.

Second, magnetic fields provide extra pressure support that prevents regions from collapsing unless their mass is higher than the *magnetic critical mass* (see also question 3.7)

$$M_\mathrm{mcm} \sim \left(\frac{5}{9G}\right)^{1/2} BR^2 = \left(\frac{5}{9\pi^2 G}\right)^{1/2}\phi,\tag{8.3}$$

where $\phi$ is the magnetic flux. Observations indicate that star-forming cores, over a wide range of size and density scales tend to have $\phi/M_\mathrm{mcm}$ values that range between 0 and $(\phi/M_\mathrm{mcm})_\mathrm{crit}$. So, magnetic fields by themselves seem unable to prevent collapse. However, magneto-hydrodynamic simulations (by Hennebelle et al. 2011) do indicate that the support they do give leads to a reduction of the number of fragments by a factor of $\sim 2$ for realistic levels of magnetization.

## Mass-accretion rate

The Kelvin-Helmholtz timescale Eq. (6.19) associated with a 25 $M_\odot$ main-sequence star of radius 9 $R_\odot$ and luminosity $10^5 L_\odot$ is only $2 \times 10^4$ yr. This is much shorter than the free-fall time Eq. (3.79) – typically a few times $10^5$ yrs, consequently such massive stars reach the zero-age main sequence before finishing their mass build-up. The star can not grow faster (in mass) than the timescale set by free-fall of gas from the envelope toward the inner regions. What kind of mass-accretion rates are needed to form a massive star on a free-fall time scale? Using Eq. (6.20) we find that for a free-fall time of $2.5 \times 10^5$ yr, an average mass-accretion rate of $10^{-4} M_\odot \mathrm{yr}^{-1}$ is needed to compile our 25 $M_\odot$ star. In principle, such high accretion rates can be accommodated if the temperature of the cloud is high, see the mass-accretion rate Eq. (6.8) in the Shu inside-out collapse model. Using this equation we find that $T \sim 160$ K is needed to achieve the high mass-accretion rate. The high temperature is associated with a high sound speed, i.e. the high $\dot{M}_{\mathrm{acc}}$ is possible because the speed at which material is accreted is high. So, what may heat the medium to such high temperatures? One heating source may be the early feedback provided by low-mass stars, whose luminosities are dominated by accretion rather than internal energy generation. Even a 1 $M_\odot$ star accreting at the relatively high rates expected in the dense regions where massive stars form could radiate strongly enough to raise the gas temperature by a factor of a few at distances of $\sim$1000 AU. Not only does this raise the minimum mass for gas to fragment by a factor of $\sim$10 (Krumholz 2006), it also increases the mass infall rate by a factor of $\sim$10. Though low-mass stars may provide considerable heating, temperatures above $\sim$100 K are only observed – though not common – in very massive star forming regions *after* one or more massive stars have formed. Therefore it is these objects that are thought to be responsible for heating the medium to such high temperatures. Perhaps this indicates that the first or first few massive stars in a cluster are formed in a completely different way, later facilitating the formation of other massive stars through the inside-out collapse mechanisms that builds lower mass stars.

Can the central object accept all of the cloud material that is falling in? Because of conservation of angular momentum the material first settles in a disk-like structure through which material spirals in and is accreted by the central object (disks are the topic of chapter 9). For a steady disk with dimensionless Shakura & Sunyaev viscosity $\alpha$ (Eq. 9.57), the accretion rate through the disk is (e.g Kratter et al. 2010)

$$\dot{M} = \frac{3\alpha a^3}{GQ} = 1.5 \times 10^{-4} \frac{\alpha}{Q} \left( \frac{T}{100\,K} \right)^{3/2} M_\odot \mathrm{yr}^{-1}, \tag{8.4}$$

where $a$ is the isothermal sound speed, and we have assumed a mean molecular weight $\mu = 2.3$. $Q$ is the Toomre parameter (see Eq. 9.63). The Shakura & Sunyaev viscosity in the disk cannot produce $\alpha > 1$, and the disk cannot be gravitationally stable if $Q < 1$, so the accretion rate through a gravitationally-stable disk is going to be $\sim 10^{-4} M_\odot \mathrm{yr}^{-1}$ only if the disk temperature is $\sim 100$ K. Notice though that if the infalling cloud material has a high enough temperature to transport $\sim 10^{-4} M_\odot \mathrm{yr}^{-1}$ one may expect that the disk can also carry

this amount, as it is not likely that the disk temperature will be substantially lower than the cloud temperature.

The process of disk heating to allow more mass to accrete has a limit: once the temperature required to stabilize the disk exceeds the dust sublimation temperature, it will not be easy to heat the disk further, and this may result in an instability so violent that the disk fragments entirely, halting further accretion (Kratter & Matzner 2006).

### Radiation pressure

The third potential problem in forming massive stars is that of radiation pressure. Material is falling in because it is subject to the gravitational acceleration $g_N = GM/r^2$. Radiation pressure is exerting an outward directed force on the in-falling material, more specifically, on the dust. This force is given by

$$g_{\rm rad} = \frac{1}{c} \int_0^\infty \kappa_\nu \, \mathcal{F}_\nu \, d\nu = \frac{\kappa}{c} \int_0^\infty \mathcal{F}_\nu \, d\nu = \frac{\kappa}{c} \mathcal{F} = \frac{\kappa}{c} \frac{L}{4\pi r^2}, \tag{8.5}$$

where

$$\kappa = \frac{\int_0^\infty \kappa_\nu \, \mathcal{F}_\nu \, d\nu}{\int_0^\infty \mathcal{F}_\nu \, d\nu} \tag{8.6}$$

is the flux mean opacity in $\rm cm^2 gr^{-1}$. $\mathcal{F}_\nu$ is the flux and $\mathcal{F}$ is the total (i.e. frequency integrated) flux. Since the radial dependence for $g_N$ and $g_{\rm rad}$ is the same, the net force will be inward only if $|g_N| > g_{\rm rad}$, or

$$\frac{L}{M} < \frac{4\pi Gc}{\kappa} = 2600 \left( \frac{\kappa}{5 \, \rm cm^2 \, g^{-1}} \right)^{-1} \frac{L_\odot}{M_\odot}. \tag{8.7}$$

All stars above $\sim 20 M_\odot$ have $L/M > 2600 \, L_\odot/M_\odot$, i.e., surpass the Eddington limit $|g_N| = g_{\rm rad}$, so the question arises: why doesn't radiation pressure expel circumstellar material and prevent stars from growing to masses substantially larger than $\sim 20 M_\odot$.

In the derivation of the infall criterium Eq. 8.7 it is implicitly assumed that the material falls in radially, i.e. that infall is a 1D phenomenon. However, it is inherently a 3D phenomenon. For instance, in a medium in which the dust (and the gas) is 'clumped', part of the photons created by the central star and disk pass through the envelope without being able to exert radiation pressure on the dust grains. In an initially turbulent cloud, Rayleigh-Taylor instabilities may produce optically thick 'fingers' of high optical depth material that can not be stopped by radiation pressure. Though this may be helpful to let matter pass through the envelope, it appears improbable that such a porous medium can exist in the inner region of the accretion disk, where matter is concentrated. Near the star however, an optically thick circumstellar disk will 'shield' the disk material from radiation pressure. In a sense, the disk beams the radiation field in the polar directions, weakening the radiation pressure on the disk but creating an outflow cavity or chimney in those polar directions through which radiation escapes.

### Ionization feedback

A fourth potential problem in forming massive stars is photo-ionization. Galactic molecular clouds generally have escape speeds below $10 \text{ km s}^{-1}$ (e.g. Heyer at al. 2009), the sound speed in $\sim 10^4 \text{ K}$ gas – i.e. in gas where hydrogen is ionized. As a result, if the gas in a star-forming region becomes ionized, the gas pressure may drive a thermal wind that will choke off accretion. This process is thought to be a major factor in limiting the star-forming efficiency of giant molecular clouds.

However, it is much less clear whether photo-ionization can limit the formation of individual massive stars. Two arguments suggest that it may not be an issue. First, Walmsley (1995) noted that the higher the mass inflow rate, the higher the density of matter around the star, and thus the higher the recombination rate and the smaller the Strömgren radius. If the radius of the ionized region is small enough that the escape speed from its outer edge is $> 10 \text{ km s}^{-1}$, then photo-ionized gas will not be able to flow away in a thermal wind and escape. For an accretion flow in free-fall onto a star, this criterium will be matched if the accretion rate satisfies

$$\dot{M}_{\text{acc}} > \frac{8\pi\mu GMQ_\circ}{2.2\alpha_{\text{B}}\ln(v_{\text{esc}}/a)} \sim 4 \times 10^{-5} \left(\frac{M}{100\, M_\odot}\right)^{1/2} \left(\frac{Q_\circ}{10^{49}\,\text{s}^{-1}}\right)^{1/2} \quad M_\odot \text{yr}^{-1}, \quad (8.8)$$

where $M$ is the stellar mass, $Q_\circ$ is the number of hydrogen ionizing photons emitted by the star per second, $\alpha_{\text{B}}$ is the case B recombination coefficient, and $v_{\text{esc}}$ is the escape speed from the stellar surface. The factor 2.2 in the denominator arrises from the assumption that He is singly ionized. The second right hand site uses $v_{\text{esc}} = 1000\,\text{km s}^{-1}$ and $a = 10\,\text{km s}^{-1}$. Thus an accretion rate of $\sim 10^{-4}\,M_\odot\text{yr}^{-1}$ is sufficient to allow continuing accretion. Given the dense, compact environments in which massive stars appear to form, such high accretion rates are entirely expected.

The second argument concerns the natural 3D geometry of the process. The accretion disk and dense filaments have very high recombination rates (because they are dense) and thus resist being photo-ionized. In low-density polar regions this self-shielding does not work. The gas will become photo-ionized and is blown out, feeding back into the interstellar medium. However, the bulk of the material is contained in dense regions, where it continues to accrete. The net effect is thus an increasing mass of the central object.

### Stellar winds

The final potential challenge for the formation of massive stars by accretion is stellar winds. Once the surface temperature of stars exceed $\sim 25\,000\,\text{K}$, they begin to develop trans-sonic radiatively driven outflows. Zero-age main sequence stars reach this temperature at a mass of $\sim 40\,M_\odot$, and stars of this mass have such short Kelvin-Helmholtz timescales that, even if they are rapidly accreting, their radii and surface temperatures during formation are likely to be close to their ZAMS values (Hosokawa & Omukai 2009). The momentum by these winds is up to about half that of the stellar radiation field, and so if the direct stellar radiation field cannot stop accretion then the momentum carried by stellar winds will not either.

However, winds might yet be important, because the wind terminal velocity (reached already close to the surface) is quite large, $\sim 1000\,\mathrm{km\,s^{-1}}$. As a result, when the winds shock against the dense accretion flow, their post-shock temperature can be $> 10^7$ K. This is well past the peak of the cooling curve (Weaver et al. 2007), and thus gas can not cool efficiently by emitting radiation. Should it become trapped, this hot gas could exert a pressure that is far greater than what would be suggested by its launch momentum. If this were to happen, it is possible that the stellar wind gas might be able to interfere with accretion. Again, we should consider possible 3D effects, i.e. clumps and filaments, that allow the (shocked) wind gas to leak out through openings in the surrounding gas rather than becoming trapped and building up a large pressure. If the medium is sufficiently porous, it is unlikely that stellar winds set significant limits on the masses to which stars can grow by accretion.

## Accretion models for massive star formation

In section 4.4 we introduced two variants of massive star formation theories, *monolithic collapse* and *competitive accretion*.

In the monolithic collapse model single molecular cloud cores of $\sim 0.1$ pc would result in the formation of a single, or perhaps a low-order multiple massive star provided that the mass of the core is large enough. The cloud core has presumably formed by condensation from a much larger molecular cloud with a mass of thousands of solar masses, but once the core has formed, the resulting collapse is little influenced by the remainder of the material in the clump. Thus low-mass stars would form from individual low-mass cores, and high-mass stars from high-mass cores. The entire mass of the core would not necessarily go into the final star; feedback processes may cause the final mass to be less than the core mass.

The second model, competitive accretion, is based on the premise that star formation is controlled not by collapse of individual cores, but by the overall collapse of a much larger region, containing initially gas with several thousand $M_\odot$ (or more). Individual fragments (cores) form at low mass and compete for the accretion of the remaining gas; also there can be interactions between the various fragments. One might expect a massive star to form in the center, since as the cloud collapses and develops a gravitational potential well, a considerable amount of gas can be funnelled toward the center. Massive (proto-)stars hence compete quite successfully for the available gas, and to some extend deprive lower mass siblings from the gas needed to grow as well.

We now introduce a third flavour of massive star formation, fragmentation-induced starvation.

Fragmentation-induced starvation is, to some extent, an added physical consideration to the two theories mentioned above. Its basic principle is that the star accretes from a rotating gravitationally unstable disk-like structure. Gravitational instabilities in such a disk lead to fragmentation of the accretion flow and the formation of companions around the central massive star. These companions subsequently compete with the initial high-mass star for the same common gas reservoir and limit its mass growth. In other words: they starve the central star of
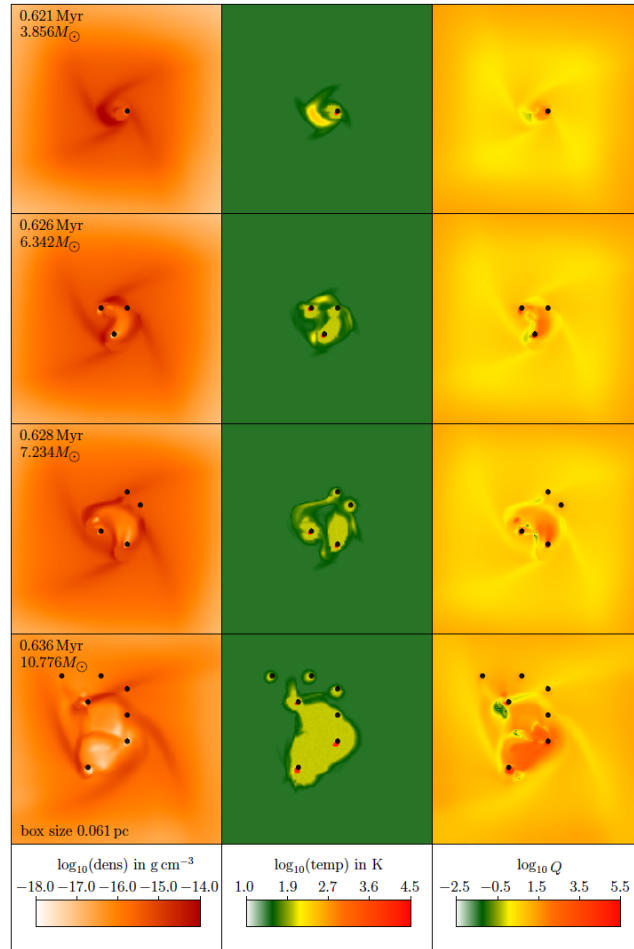
Figure 8.2: *The initial phases of disk instability in the fragmentation-induced starvation scenario. The panels show slices of density, temperature, and Toomre Q-parameter (see Eq. 9.63) in the disk plane for four different times. Each frame shows the simulation time and the mass of the most massive star in the simulation. The black dots indicate the positions of sink particles. The stellar radiation initially heats up the disk locally, which enhances the stability. The dense filaments, however, shield the radiation, and the cold material within and behind the filaments becomes unstable again. From: Peters et al. 2010, ApJ 725, 134.*

in-falling material. This is in contrast to the competitive accretion model, where the massive stars are never hindered in growth by the low-mass stars in the cluster.

Figure 8.2 shows results from a 3D radiation-hydrodynamical simulation of massive star formation that includes heating by ionising and non-ionising radiation by Peters et al. (2010, ApJ 725, 134) for a volume of size 0.24 pc × 0.24 pc × 0.015 pc. One 'observes' the continuous formation of new sink particles (i.e. proto-stellar companions) in an expanding region around the central star. The heating by stellar radiation can suppress instability locally, but shielding

Figure 8.3: Left. *Radius of sink formation as a function of time for the fragmentation-induced starvation scenario. Sink formation gradually occurs at larger disk radii. The accretion heating by the first s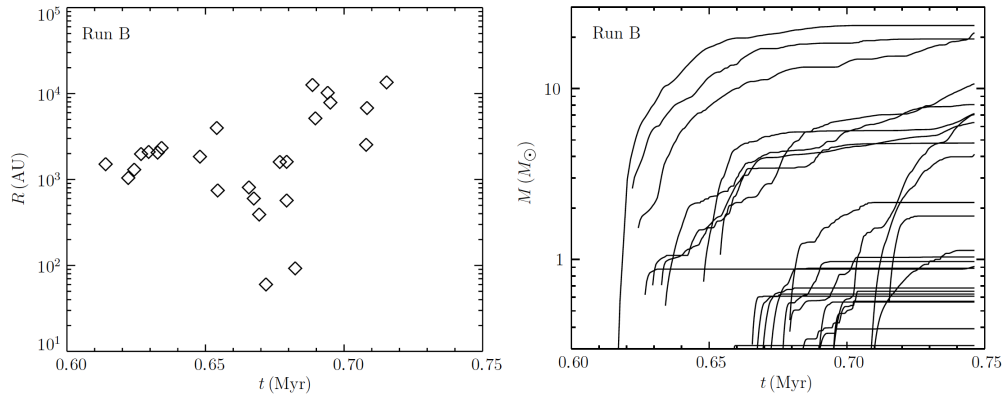tars to form suppresses sink formation at small disk radii until relatively late in the system evolution. Left. Individual accretion history of the sink sources, showing the stellar masses as a function of time for all sink particles. From: Peters et al. 2010, ApJ 725, 134.*

by the dense filaments prevents the whole disk from becoming stable and restricts the heating to small regions near the center of the disk that are surrounded by filaments. This shielding makes it possible for star formation to progress radially outward despite heating by the stars. The disk remains sufficiently cool at the inner edge for gravitational instability to set in and star formation to proceed inside-out in the disk plane. So, the effect of the filamentary structures in the disk is twofold: (1) they are so dense that they render the disk unstable locally, and (2) because of their high density, they can effectively shield the radiation from the stellar configuration (the primary star and its companions) near the center of the disk, so that radiative heating does not stabilise the outer-parts of the disk.

In the first frame at 0.621 Myr the disk contains about $35\,M_\odot$ of gas. The total disk mass remains about constant while star formation proceeds radially outward. To illustrate the tendency of star formation to occur at increasingly larger disk radii, the left panel of Fig. 8.3 shows the disk radii at which new sink particles form as a function of time. Accretion heating in the initial phase suppresses sink formation at small radii. The companions slowlffy spiral outward with time, so that at $\tau \sim 0.67\,$Myr their radiation can be shielded by filaments in the disk. Within these filaments, the gas then cools until local collapse sets in and two sink particles form near the center of the disk. Once the filament has dissolved, the gas heats up again and no further sink particles form in the inner disk region. Eventually, companions can be found up to $\sim 20\,000$ AU (or 0.1 pc) from the first source. The simulation produces a total of 25 sinks for a combined mass of $125.6\,M_\odot$ of which the most massive star has a mass $M_{\max} = 23.4\,M_\odot$. The individual accretion history of the sink sources is shown in the right panel of Fig. 8.3.

## 8.2  The formation of massive stars by collision

Another quite drastic scenario is that gravitational collapse indeed does not lead to the formation of objects more massive than some ten solar masses, and that more massive objects are the result of coalescence of low- and intermediate-mass stars. Some support for this idea could come from the observation that massive stars do originate in fairly dense clusters and that stars are seen to 'run away' from young clusters. The latter may be the result from dynamical ejection when two stars experience a close encounter[1].

The problem with coalescence is that it requires very high stellar densities to be effective. In a system containing stars of mass $M_\star$ and radius $R_\star$ that has a stellar density $n_\star$ and where the velocity distribution of the objects is according to Maxwell with a 1D velocity dispersion $\sigma$, the collision time is given by (e.g. Binney & Tremaine 1987, equation 8-125 or Freitag et al. 2006, MNRAS 368, 121, equation 3)

$$\tau_{\mathrm{coll}} \simeq 2.1 \times 10^{12} \left( \frac{10^6 \, \mathrm{pc}^{-3}}{n_\star} \right) \left( \frac{\sigma}{30 \, \mathrm{km \, s^{-1}}} \right) \left( \frac{R_\odot}{R_\star} \right) \left( \frac{M_\odot}{M_\star} \right) \quad \mathrm{yr.} \qquad (8.9)$$

Inserting values typical for the central cluster Radcliffe 136 of the 30 Doradus region (see figure 8.1), where $n \sim 1.5 \times 10^4$ (Selman & Melnick 2013, A&A 552, A94, adopting a typical core cluster size of 0.4 pc) and $\sigma \sim 5 \, \mathrm{km \, s^{-1}}$ (Hénault-Brunet et al. 2012, A&A 545, L1), and pretending all stars have a mass $M_\star = 1 \, M_\odot$ and radius $R_\star = 1 \, R_\odot$, yields $\tau_{\mathrm{coll}} \simeq 2.5 \times 10^{13}$ yr. At the present-day evolutionary state of the cluster collisions are thus negligible.

### Gas accretion-driven collision models

The situation may, however, have been quite different early on in the cluster evolution. Right now, most of the gas that did not end up in stars has been blown out. This may have expanded the size of the core by a factor of ten, implying a $10^3$ shorter collision time during the early phases of formation. The velocity dispersion will, however, have been higher, with typical values of tens of $\mathrm{km \, s^{-1}}$. Though these considerations help to bring down the collision time, the collision efficiency would be much higher if the merging would happen during an earlier phase of evolution, i.e. during the stage when the future stars are still clumps (of size $\sim$ 10 AU $\simeq$ 2000 $R_\odot$) or are surrounded by dense proto-stellar disks (of size $\sim$ 100 AU $\simeq$ 20000 $R_\odot$). Taking all these considerations into account, the collision time may be reduced to $\sim 0.5 - 5$ Myr. This is still a long time interval, as for stellar mergers to be efficient in making massive stars, the timescale $\tau_{\mathrm{coll}}$ should be of the order of $0.1 - 1$ Myr. Perhaps a further reduction of $\tau_{\mathrm{coll}}$ can be achieved if the competition for gas in the context of competitive accretion is also considered. This would allow massive cores or proto-stars to grow faster,

---

[1]In older clusters stars may be ejected from binary systems in which the initially most massive component experiences its supernova explosion.

becoming more massive prior to merging events. The mass scaling in Eq. 8.9 may then imply a reduction of the collision time by another factor a few.

So perhaps a combination of merging, starting already in the early core phase, and competitive accretion may be the road towards massive star formation. It seems likely to assume that near the end of the merger phase some, or many, spiraling, massive stars do not complete their merger, but remain in tight orbits. Indeed, it has been established that the fraction of close binaries (with periods less than 1500 days) among massive stars is high – about 70 percent – and that many of these binaries are very compact (Sana et al. 2012, Science 337, 444). About 1/3th of these binaries orbit each other in less than six days when arriving on the main sequence. These periods are so short that these stars will merge during their lifetime as a result of stellar evolution. Indeed, the formation process of massive stars may start in the cloud core phase and continue all the way into the main sequence phase.

## 8.3 Formation of the First Stars

The discussion in this section closely follows and summarizes the excellent and accessible review by Bromm et al. (2009, Nature 459, 49).

The cold dark-matter (CDM) model of cosmic evolution posits that structure grew hierarchically, such that objects of small size formed first and then merged to form increasingly larger systems. Within this model dark-matter 'mini-haloes' formed first through the action of gravity alone, probably a few hundred million years after the Big Bang (corresponding to redshifts $z \approx 20-30$) in an era referred to as the 'dark ages'. In simulations of structure formation these dark matter mini-haloes, with virial temperature $T \sim 1\,000$ K and mass $M \sim 10^6 M_\odot$, are identified as the sites where the first stars formed. They correspond to $3\sigma$ to $4\sigma$ peaks in the cosmic density field. Such high-density peaks are expected to be strongly clustered, suggesting that feedback effects from the first stars are important in determining the fate of the surrounding primordial gas clouds. It is for this reason that two classes of population III stars, i.e. stars composed of no elements heavier than helium (ignoring the small amount of lithium produced in baryogenesis), are defined. First generation population III.1 stars form from initial conditions determined entirely by cosmological parameters, and second generation population III.2 stars, which originate from material that was influenced by nearby population III.1 stars.

For star formation to begin a sufficient amount of gas must accumulate in a dark matter minihalo. This gas was essentially atomic in nature, with only small traces of molecular hydrogen. These molecules can be produced via a sequence of reactions, starting with the formation of the hydrogen anion $H^-$ by radiative association

$$H + e^- \rightarrow H^- + h\nu. \tag{8.10}$$

This is followed by the formation of $H_2$ by associative detachment

$$H^- + H \rightarrow H_2 + e^- + E_{kin}. \tag{8.11}$$

This channel toward $H_2$ production is, however, not efficient and a fractional abundance of $H_2$ of $10^{-5}$ is expected. Hydrogen or helium atoms can not efficiently cool the primordial gas by emitting radiation because they have excitation energies that are high, the primary cooling channel being collisional excitation of the hydrogen $1s \rightarrow 2p$ transition followed by an emission of a Ly$\alpha$ photon. The minimum temperature that can be reached in this way is $\sim 10^4$ K. The lowest accessible rotational energy level of $H_2$ has an excitation energy of $510$ K, implying that no cooling mechanism is available to lower the temperature of the gas below a few hundred Kelvin. This is a fundamental difference with gas clouds in the local universe, where dust particles and molecules such as CO can cool the gas to temperature of $\sim 10$ K.

The minimum mass at the onset of collapse is determined by the Jeans mass. Rewriting Eq. 3.63 to suit our needs, we get

$$M_{\rm J} \sim 1000 \left( \frac{T}{200\,\mathrm{K}} \right)^{3/2} \left( \frac{10^4}{n} \right)^{1/2} M_\odot, \tag{8.12}$$

where we have adopted $\mu = 1.3$. Note that the critical density of the para-$H_2$ ($J = 2 - 0$) transition (of $510$ K) – above which cooling becomes less efficient – is only $10$ cm$^{-3}$, suggesting Jeans masses of thousands of solar masses. Whether the collapse leads to the formation of a single object or a multiple systems is presently unclear. Simulations based on self-consistent cosmological initial conditions seem to favor a single object, however those starting from non-cosmological initial conditions have yielded multiple cloud cores. Simulations assuming a high initial degree of spin lead to disk formation and its subsequent break-up.

Although we can fairly accurately estimate the mass triggering the first runaway collapse, it does not really tell us what the mass of the forming star(s) will be. The final mass is determined by both the mass of the cloud of which it forms and by the efficiency of feedback processes that occur during the evolution of the proto-star. These feedback processes may differ in several important ways in massive primordial stars from those of massive stars in the present-day universe. First, primordial gas does not contain dust and therefore radiative forces on the gas are much weaker. Second, it is generally assumed that magnetic fields are not important in primordial gas. Such fields in the present-day universe help to drive powerful outflows and remove angular momentum (see section 8.1). Third, primordial stars are much hotter than contemporary stars of the same mass, resulting in significantly greater ionizing luminosities.

Once the first massive proto-star is formed, be it in isolation or as part of a multiple system, its far-ultraviolet radiation in the Lyman-Werner bands (in the range of 11.2 to 13.6 eV) will photo-dissociate all the $H_2$ in the parent gas cloud. This reduces the cooling rate, but simulations do not indicate that this stops the accretion. Radiation pressure (in the Lyman-$\alpha$ line) can reverse the infall in the polar regions once the proto-star has grown to 20–30 $M_\odot$, but cannot significantly reduce the accretion rate. Once the proto-star reaches 50–100 $M_\odot$, the expansion of the H II region can reduce the accretion rate, but accretion can continue in the equatorial plane. It is ultimately photo evaporation-driven mass loss from the disk that is thought to halt
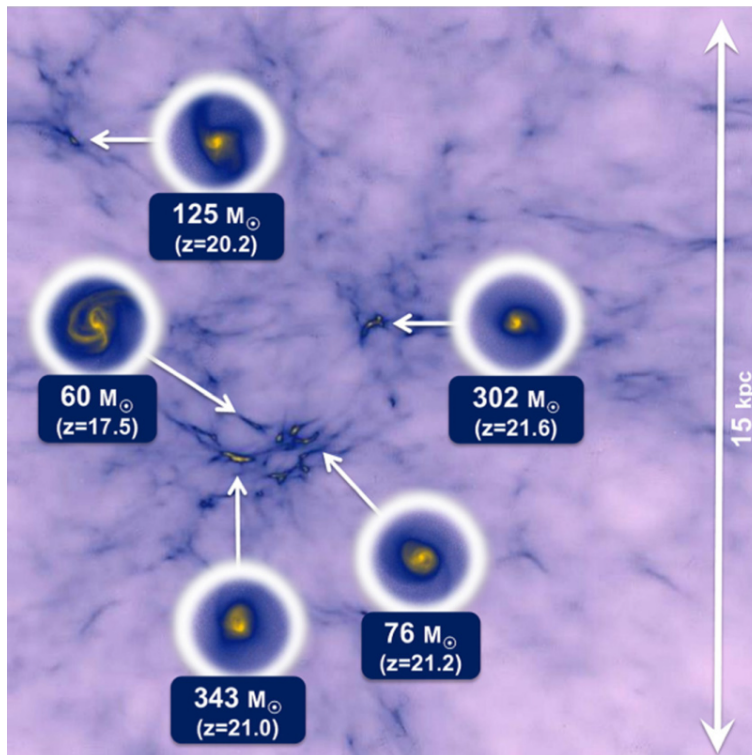
Figure 8.4: *Projected gas density distribution at redshift $z = 25$ in a radiation hydrodynamical simulation of primordial star formation in a cosmological context. Five primordial star-forming clouds in a cube of 15 kpc on a side are shown. The circles provide a zoom-in to the central 1pc region of the clouds at the respective formation epochs. The redshifts of these epochs and the masses of the first stars formed in these clouds are given in the blue boxes. From: Hirano et al. 2014, ApJ 781, 60.*

the accretion and fix the final mass of the stars. For reasonable conditions simulations favor final masses of 10–1000 $M_\odot$, depending on the entropy and angular momentum of the natal cloud and the mass-accretion rate (e.g. Hirano et al. 2014; see also Fig. 8.4).

## The role of the first stars in the formation of the first galaxies

Cosmological simulations aimed at producing the first galaxies quite generally show that the first stars form before the first galaxies do, and that feedback effects from these stars play a key role in determining the initial conditions for the formation of the first galaxies and of subsequent star formation in these primordial galaxies. Lyman-Werner radiation and hydrogen-ionizing radiation emitted by the first stars may have heated and pressurized the ambient medium (on kpc scales), reducing the density, and overall delaying the formation of the second generation of stars by as much a 100 Myr. If radiation driven winds of population III stars are weak and mass eruptions are absent, such that the pre-supernova masses remain high, stars with initial masses in the range $25 \lesssim M \lesssim 140\,M_\odot$ and $M \gtrsim 260\,M_\odot$, end their lives

by collapsing into black holes with relatively little ejection of heavy elements. Population III stars in the range 140–260 $M_\odot$ explode as pair-instability supernovae, leaving no compact remnant. The nucleosynthetic yields of such explosions, defined as the heavy-element mass fraction, may reach 0.5. If this leads to metallicities in the surrounding gas in excess of a critical value of $Z_{\rm crit} \sim 10^{-4}\, Z_\odot$ it may be enough to increase the cooling efficiency such that low-mass (population II) stars may form in a next event of star formation.

It is possible that at least one primordial star has formed in a region that is destined to eventually become a first galaxy. However, current simulations do not yet fully reach the required resolution and physical realism to test this. At the present time, a promising idea to pursue seems to be an atomic cooling halo – i.e. a halo in which the Lyman-Werner ultraviolet flux from population III.1 stars in the surroundings exceeds a critical threshold, effectively destroying $H_2$ – with $\sim 10^8\, M_\odot$ and a virial temperature greater than $\sim 10^4$ K. The dynamical time for a first-galaxy halo to assemble is estimated at $\sim 10^8$ yr, i.e. of the same order as the delay time between the first and second generation of star formation. The time it takes for the chemically enriched material ejected by pair-creation population III.1 supernova to disseminate and be re-incorporated in a new (population III.2) star or galaxy formation event may also reach up to $10^8$ yr. Recent cosmological simulations indicate that – in marked contrast to the situation in dark matter mini-haloes – star formation in the first galaxies was strongly influenced by gravitationally driven supersonic turbulence generated during the virialization process. This, together with the chemical enrichment, may have lead to the formation of the first stellar clusters. It remains however an open question as to whether the first galaxies were the sites where globular clusters formed, which are the oldest star clusters known.

Discoveries of quasars at redshifts of $z \sim 6-7$ that are powered by black holes with masses of $\sim 10^9\, M_\odot$ show that supermassive black holes (SMBH) already came into existence within 1 Gyr after the Big Bang. The current paradigm implies that these objects have grown, starting from a much smaller seed, via matter accretion and, to a lesser extent by merging with other compact objects (see e.g. Natarajan 2011). To allow growth to the observed masses, the initial seeds must have been $\sim$400 $M_\odot$ or more. Such objects may have been formed in supernova explosions in the high-mass tail of the first stars.

An alternative route toward the formation of SMBH is that of 'direct collapse black hole (DCBH) formation' by gas gravitational collapse – proceeding at rates of $\gtrsim 0.1 - 1\, M_\odot {\rm yr}^{-1}$, i.e. $\sim$100 times larger than for standard metal-free star formation – in dark matter haloes with virial temperatures $T \gtrsim 10^4$ K (e.g. Dijkstra et al. 2014). Given the strong temperature sensitivity of the atomic cooling process the gas collapses almost isothermally, thermostatting at $T \sim 8\,000$ K. Under these conditions, it seems that gas fragmentation into sub-clumps is almost completely inhibited and collapse proceeds to very high densities unimpeded. To conclude, DCBH formation requires the following conditions on the host haloes: (1) virial temperature $> 10^4$ K to ensure high accretion rates allowed by atomic cooling; (2) gas metallicities $Z < Z_{\rm crit}$ to prevent fragmentation into clumps induced by heavy elements and dust cooling; (3) exposure to a Lyman-Werner ultraviolet radiation field from a population of first stars that is strong enough to strongly depress $H_2$ abundances.

We conclude that though much is still unclear as to the mechanism of galaxy formation, the role of the first stars in the shaping of these systems seems pivotal.

# Properties of circum-stellar disks

The irrefutable confirmation that disks are a common by-product of star formation dates back to the late 1980's and early 1990's, when the first millimeter images (Beckwith et al. 1986, Sargent & Beckwith 1987, Rodriguez et al. 1992) and spectroscopy (e.g. Koerner et al. 1993) of such disks were taken. The millimeter dust emission revealed extended structures around young stars and the double-peaked line profiles were proof of a regular Keplerian rotation pattern (see Fig. 9.1). The launch of the Hubble Space Telescope in 1990 opened a new window for research on proto-planetary disks. The high spatial resolution from space enabled the detailed imaging of the star-forming Orion nebula at a distance of 450 pc (O'Dell et al. 1993, ApJ 410, 696; O'Dell 1993, Science 259, 3, McCaughrean & O'Dell 1996, AJ 111, 1977). These images (see Fig. 9.4) were taken with the Wide Field Camera in several optical narrow band filters such as H$\alpha$, [O III], [O I] and [S II]. This data show not only that such proto-planetary disks are ubiquitous around newly formed stars – in $\sim$50% of them a disk is detected – but reveals also the impact of disk irradiation and erosion by nearby hot O and B stars. Proto-planetary disks are now routinely resolved from the ground with the Atacama Large Millimeter/submillimeter Array (ALMA) that began scientific observations in the second half of 2011.

This chapter will focus on the formation of the disk, its radial and vertical structure, the transport of angular momentum — a prerequisite for accretion of material onto the star — and jets & outflows. Figure 9.2 shows a kaleidoscopic overview of the current paradigm of the structure of a typical proto-planetary disk. Many of the basics of accretion disks are covered in the review article by Pringle (1981). Here we closely follow and use excerpts from the excellent textbook by Armitage (2010), 'Astrophysics of Planet Formation'.

## Disk formation

Section 7.1 describes how a flat disk forms during the collapse of a rotating cloud. Since the simple collapse problem outlined there is spherically symmetric and the rotation itself axisymmetric, the trajectories of infalling material from the top are mirrored at the equato-
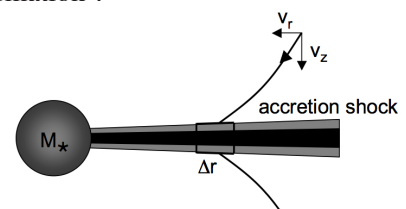


Figure 9.3: *Schematic view of disk formation during the collapse of a rotating spherical cloud.*
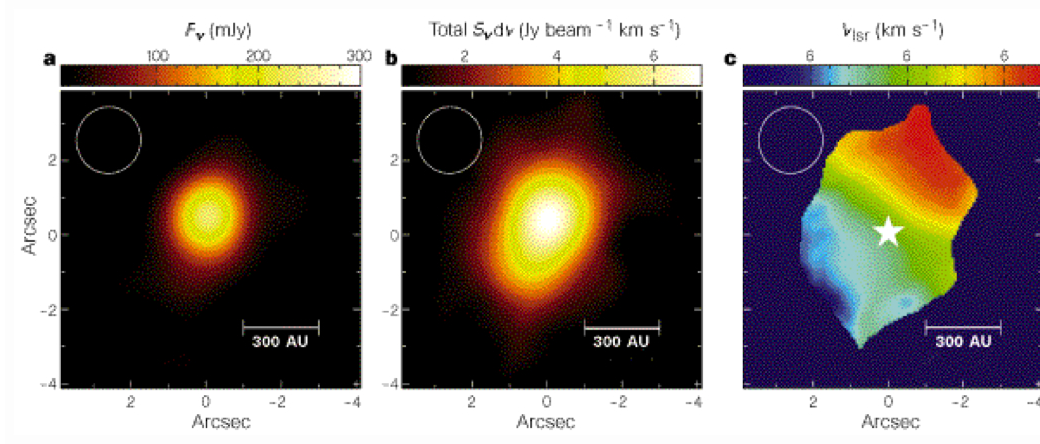
Figure 9.1: *Images of the circum-stellar disk around MWC 480. The star symbol in the right panel indicates the position of the central source. In each panel, the full-width at half-maximum of the synthesized beam is shown at top left, and offsets in arcseconds from the stellar position are indicated along the vertical and horizontal axes. Left panel: Thermal continuum from dust grains, at $\lambda = 1.3$ mm. The flux density $\mathcal{F}_\nu$ scale is shown at the top of the panel. Middle panel: CO(2-1) emission, integrated across the full velocity range (+2.4 to +7.6 $\mathrm{km\,s^{-1}}$; the stellar systemic velocity is at $\sim 5.1$ $\mathrm{km\,s^{-1}}$). The scale at the top of the panel provides the integrated intensities ($S_\nu \, d\nu$). Right panel: Intensity-weighted mean gas velocities at each spatial point of the structure shown in the middle panel. The stellar systemic velocity is represented by green. Blue-shifted (approaching) and red-shifted (receding) velocity components are shown as blue and red, respectively. From: Mannings et al. (1997).*

rial plane. The in-falling fluid elements, which can well be represented by parabolic trajectories, thus collide in the equatorial plane with matter coming from the other side of the cloud. An accretion shock forms where the material collides and this accretion shock extends over the entire centrifugal radius $r_{\mathrm{cf}}$. If the excess heat energy from the shock can be efficiently radiated away, a thin disk structure forms in the equatorial plane (Fig. 9.3).

In the radial direction, the disk is supported against gravity by its rotation. Gas pressure only plays a minor role. To first order we can thus approximate the orbital velocity of the gas by the *Keplerian orbital speed* $v_\phi$, that can be found by equating the centripetal force $F_{\mathrm{cp}} = m \, v^2/r$ to the gravitational force $F_{\mathrm{gr}} = G \, m \, M_*/r^2$. This yields

$$v_\phi = v_{\mathrm{K}} = \Omega_{\mathrm{K}} \, r = \sqrt{\frac{GM_*}{r}} = 29.8 \left(\frac{M_*}{M_\odot}\right)^{1/2} \left(\frac{r}{1\,\mathrm{AU}}\right)^{-1/2} \quad \mathrm{km/s}, \qquad (9.1)$$

where $\Omega_{\mathrm{K}}$ is the Kepler frequency

$$\Omega_{\mathrm{K}} = \sqrt{\frac{GM_*}{r^3}}. \qquad (9.2)$$

In the shock and subsequent cooling process, the component of the infalling velocity perpendicular to the disk mid-plane $v_z$ is dissipated, so that the material within the disk is left with
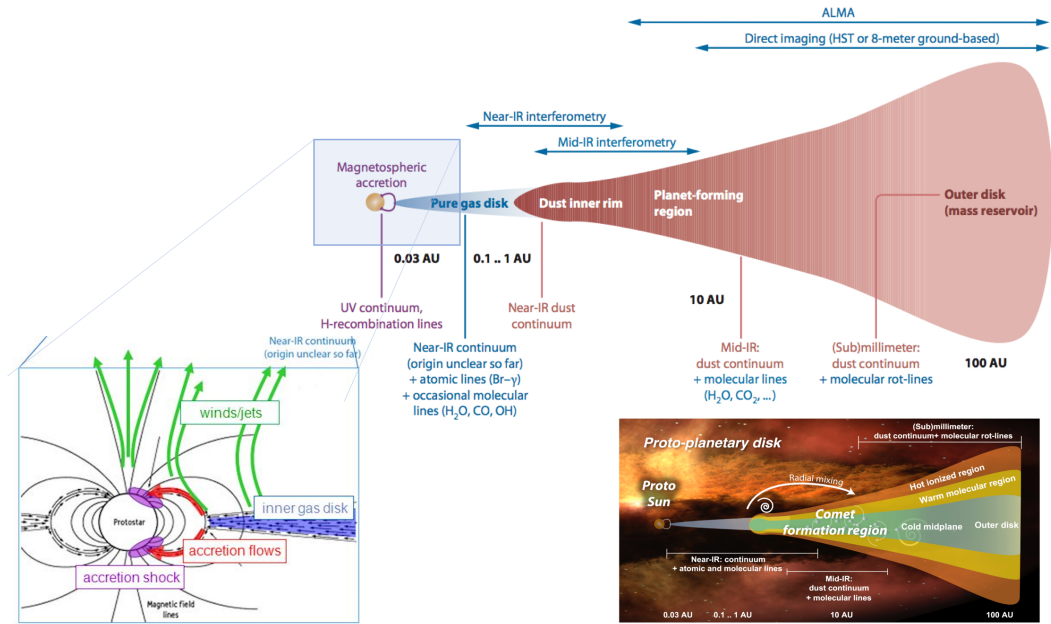
Figure 9.2: *Artist impression depicting the current paradigm of proto-planetary disk structure. The conditions in these disks span a wide range in temperatures and densities. They are therefore studied with a suite of instrumentation, probing cold and warm dust, cold molecular gas, warm atomic gas, and ionized gas close to the central star and in a thin surface layer of the disk. At a distance of 100 pc, a disk of 100 AU measures 0.01" = 10 milli-arcsec in the plane of the sky. Adapted from Dullemond & Monnier 2010, Punzi 2014, and Antoniucci 2015.*

the in-plane component of its in-falling velocity $v_r$ (see Fig. 9.3). The point at which a parcel of in-falling material hits the disk depends entirely on its initial angular momentum, so its initial position within the cloud. It is thus to be expected that some additional mixing and angular momentum transport within the disk is needed until its angular momentum and mass is redistributed in such a way that we end up with a disk in nearly Keplerian rotation. Realize too, for material to be transported inward and accreted by the star, it must lose angular momentum. We may easily fathom this: for a geometrically thin disk the angular velocity of the gas is essentially the Keplerian velocity Eq. 9.1 (but see Eq. 10.19), hence the specific angular momentum

$$j = r\,v_\phi = \sqrt{GM_* r},\tag{9.3}$$

is an increasing function of radius. For gas in the disk to flow toward the central star, it thus needs to lose angular momentum. If accretion occurs, the disk must thus evolve. In the next section, we derive the equation that describes this evolution.
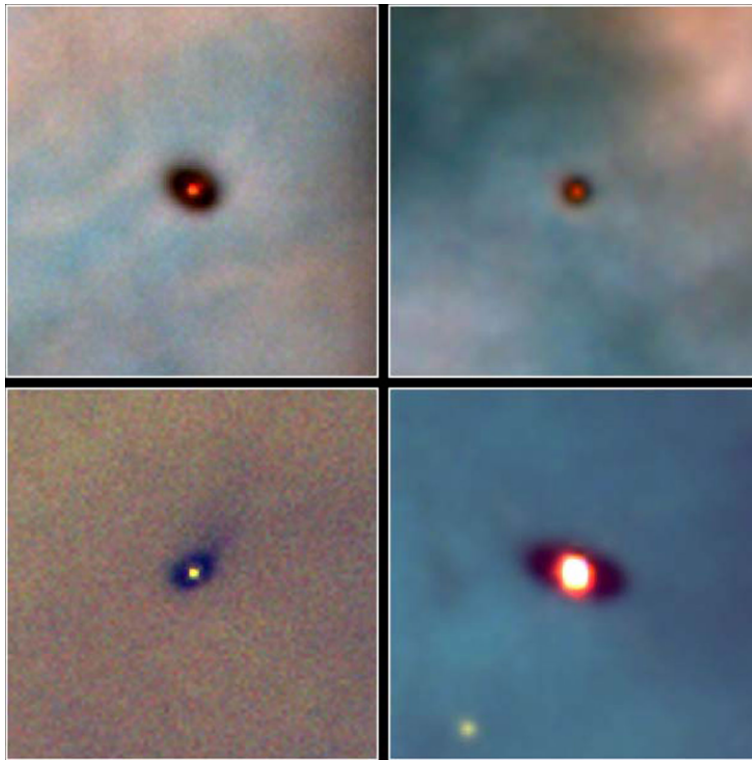
Figure 9.4: *Wide Field Camera 2 color composite images of proto-planetary disks in the Orion Nebula. Blue colors indicate O III $\lambda502$ nm emission; green colors H$\alpha$ emission, and red colors N II $\lambda658$ nm. The disks appear dark because the dust in these disks absorbs the bright nebular gas emission by the background nebular gas. From: McCaughrean & O'Dell 1996, AJ 111,5. Image PRC95-45b (1995), ST Sci OPO, from HST archive.*

## 9.1  Disk structure

We consider in the following an axisymmetric disk and assume that the surface density can be written in the general form of $\Sigma(r, t)$, which is an integral of the mass density $\rho(r, z)$ over the vertical disk height $z$

$$\Sigma(r, t) = \frac{1}{2\pi} \int_0^{2\pi} \int_{-\infty}^{+\infty} \rho(r, z, t)\, dz d\phi = \int_{-\infty}^{\infty} \rho(r, z, t) dz. \qquad (9.4)$$

In deriving the general equations of disk structure, we follow the motion of an annulus of the disk and use the vertically integrated form in all equations, i.e. we employ the surface density $\Sigma(r)$ in gr cm$^{-2}$ instead of the physical density $\rho(r)$. Matter can flow into the annulus and out of it with a radial velocity $v_r$.

## Surface density evolution of a thin disk

Because of the chosen geometry it is most logical to express the relevant equations in cylindrical polar coordinates. The equation of mass conservation (3.10) is then given by

$$\frac{\partial \rho}{\partial t} + \frac{1}{r}\frac{\partial}{\partial r}\left(r\,\rho\,v_r\right) + \frac{1}{r}\frac{\partial}{\partial \phi}\left(\rho\,v_\phi\right) + \frac{\partial}{\partial z}\left(\rho\,v_z\right) = 0. \tag{9.5}$$

Integrating with respect to $\phi$ and $z$, and realizing that because $\partial v_\phi/\partial\phi$ and $\partial\rho/\partial\phi$ vanish in an axisymmetric disk the third term is zero and because of an assumed absence of velocities perpendicular to the disk this is the case for the fourth term as well, yields

$$r\,\frac{\partial \Sigma}{\partial t} + \frac{\partial}{\partial r}\left(r\,\Sigma\,v_r\right) = 0. \tag{9.6}$$

Similarly, conservation of angular momentum for the annulus yields the equation

$$r\frac{\partial}{\partial t}\left(r^2\,\Omega\,\Sigma\right) + \frac{\partial}{\partial r}\left(r^2\,\Omega\cdot r\,\Sigma\,v_r\right) = \frac{1}{2\pi}\frac{\partial G}{\partial r}. \tag{9.7}$$

This derivation is more elaborate as it involves the introduction of a stress tensor to describe, for instance, turbulence or magnetic fields. We leave this to the endevouring reader. The two velocity components are $v_r$ and $v_\phi = \Omega\,r$. The angular momentum of a thin annulus of the disk is $2\pi r\,\Delta r\,\Sigma\,r^2\,\Omega$, where $\Delta r$ is the width of the annulus. The rate of change of angular momentum of this thin annulus is determined by the change in surface density due to radial flows (the second term on the left-hand side) and by the *difference* in the torque $G$ exerted on the thin annulus by stresses at its inner and outer edges.

We briefly revisit here the concept of a torque $G$. A force that acts on the center of mass of a body will cause a linear motion. If the same force acts on a point off the mass center, the body will start to rotate as well. The torque is then the product of the force $F$ and the lever, i.e. the distance from the mass center $r$

$$G = rF\sin\theta, \tag{9.8}$$

where $\theta$ is the angle between the force vector and the lever vector. The accretion disk is not rotating as a rigid body and so we can apply the concept of torque on each annulus in the disk. The neighboring annulus will exert a force on it proportional to the rate of shear $\{dv_\phi/dr\}_{\text{shear}}$ between the two annuli. From

$$\frac{\mathrm{d}v_\phi}{\mathrm{d}r} = \frac{\mathrm{d}(\Omega\,r)}{\mathrm{d}r} = r\,\frac{\mathrm{d}\Omega}{\mathrm{d}r} + \Omega = \left\{\frac{\mathrm{d}v_\phi}{\mathrm{d}r}\right\}_{\text{shear}} + \left\{\frac{\mathrm{d}v_\phi}{\mathrm{d}r}\right\}_{\text{rigid}} \tag{9.9}$$

where $\{dv_\phi/dr\}_{\text{rigid}}$ is the velocity gradient required to maintain rigid rotation, we find that the force associated with the rate of shearing is proportional to $r\,d\Omega/dr$. It is called the shear force.

Without describing yet the details of the viscosity or angular momentum transport, we note that for a geometrically thin disk

$$G = 2\pi\,r\cdot\nu\,\Sigma\,r\frac{\mathrm{d}\Omega}{\mathrm{d}r}\cdot r = 2\pi r^3\,\nu\,\Sigma\frac{\mathrm{d}\Omega}{\mathrm{d}r}. \tag{9.10}$$

Here, $\nu$ is the kinematic viscosity in $\text{cm}^2\,\text{s}^{-1}$ and $G$ is the torque exerted on an annulus in the disk. It is the product of the circumference, the viscous force per unit length ($\nu\,\Sigma\,r\,d\Omega/dr$) and the lever $r$.

Using the expression for the torque $G$, we can rewrite Eq. (9.7) as

$$\frac{\partial}{\partial t}\left(r^2\,\Omega\,\Sigma\right) + \frac{1}{r}\frac{\partial}{\partial r}\left(r^2\,\Omega\cdot r\,\Sigma\,v_r\right) = \frac{1}{r}\frac{\partial}{\partial r}\left(\nu\,\Sigma\,r^3\,\frac{d\Omega}{dr}\right). \tag{9.11}$$

We now assume that the orbital frequency equals the Kepler frequency. In that case, the specific orbital angular momentum $j = r^2\,\Omega$ is a function of radius only, hence $\partial\left(r^2\,\Omega\right)/\partial t = 0$. Using the continuity equation (Eq.9.6) to eliminate the time dependance, the above result simplifies to

$$\Sigma\,v_r\frac{\partial(r^2\,\Omega)}{\partial r} = \frac{1}{r}\frac{\partial}{\partial r}\left(\nu\,\Sigma\,r^3\,\frac{d\Omega}{dr}\right). \tag{9.12}$$

For $\Omega = \Omega_{\mathrm{K}}$ (see Eq. 9.2) the radial partial derivative can be worked out. This yields an expression for the radial velocity in the disk

$$v_r = -\frac{3}{\Sigma\,\sqrt{r}}\frac{\partial}{\partial r}\left(\nu\,\Sigma\,\sqrt{r}\right). \tag{9.13}$$

Alternatively, we can insert Eq. (9.12) into Eq. (9.6) to eliminate $v_r$ and obtain

$$\frac{\partial\Sigma}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left[\left(\frac{d}{dr}(r^2\,\Omega)\right)^{-1}\frac{\partial}{\partial r}\left(\nu\,\Sigma\,r^3\left(-\frac{d\Omega}{dr}\right)\right)\right]. \tag{9.14}$$

Using our assumption of Keplerian rotation, we can simplify this to

$$\frac{\partial\Sigma}{\partial t} = \frac{3}{r}\frac{\partial}{\partial r}\left[\sqrt{r}\frac{\partial}{\partial r}\left(\nu\,\Sigma\,\sqrt{r}\right)\right]. \tag{9.15}$$

This is the evolution equation for the surface density of a geometrically thin Keplerian accretion disk under the action of internal angular momentum transport. The evolution is a diffusive partial differential equation for the surface density $\Sigma(r)$. It is linear if the viscosity $\nu$ is not itself a function of $\Sigma$. The equation can also be derived directly from the Navier-Stokes equation for a viscous fluid in cylindrical polar coordinates.

## Radial surface density structure of a steady-state accretion disk

In steady state, i.e. $\partial/\partial t = 0$, the disk structure follows from radial and angular momentum conservation and the assumption that the vertical component of gravity from the star is balanced by the vertical gas pressure gradient. The latter implies hydrostatic equilibrium in the vertical direction. We study a disk annulus at a distance $r$ from the star. Material can flow

into and out of the annulus with the velocity $v_r$ in the radial direction. If magnetic fields and viscosity are negligible, the radial momentum equation for a steady-state flow is (see Eq. 3.23)

$$v_r \frac{\partial v_r}{\partial r} - \frac{v_\phi^2}{r} + \frac{1}{\rho} \frac{\partial p}{\partial r} + \frac{G M_\star}{r^2} = 0, \tag{9.16}$$

where $v_r$ is the radial velocity of the gas, $v_\phi$ is the circular velocity, and $M_\star$ is the mass of the central star. The four terms in the above equation arise from radial mass flow, centrifugal force, force due to gas pressure gradients, and gravity. Since pressure typically decreases with increasing radius, the third term is nearly always negative. Effectively, gas pressure resists the gravitational force, resulting in gas rotating at sub-Keplerian orbital velocities.

Following Eq. 9.11, the angular momentum conservation equation becomes

$$\frac{\partial}{\partial r} \left( r^2 \, \Omega \cdot r \, \Sigma \, v_r \right) = \frac{\partial}{\partial r} \left( \nu \, \Sigma \, r^3 \, \frac{d\Omega}{dr} \right), \tag{9.17}$$

where the left-hand side is the radial change in angular momentum and the right-hand side arises from the viscous torques. Integrating this equation and multiplying by $2\pi$ yields

$$2\pi \, r \, \Sigma \, v_r \cdot r^2 \, \Omega = 2\pi \, r^3 \, \nu \, \Sigma \, \frac{d\Omega}{dr} + \text{constant} \tag{9.18}$$

Noting that the mass accretion rate $\dot{M} = -2\pi \, r \, \Sigma \, v_r$ we can write this equation in the form

$$-\dot{M} \cdot r^2 \, \Omega = 2\pi \, r^3 \, \nu \, \Sigma \, \frac{d\Omega}{dr} + \text{constant}, \tag{9.19}$$

where the constant of integration, which has the form of an angular momentum flux, remains to be determined. This constant may be specified by introducing the notion that there must be a location in the disk where the rate of change of the angular frequency is zero, i.e., where $d\Omega/dr = 0$ (see Fig. 9.5). If we start far out in the disk and move towards the central star, the Keplerian frequency Eq. (9.2) will increase. If the disk would extend to the stellar surface and the star would be spinning at break-up speed, indeed $\Omega$ (from outside in) would continue to rise. Young stellar objects are, however, rarely observed to spin at break-up speed, hence $\Omega$ must peak (where $d\Omega/dr = 0$) and turn-over to smoothly connect to the stellar rotation rate. At the location where $d\Omega/dr = 0$, in most cases close to the stellar surface, the viscous stress vanishes, and the constant is simply equal to the flux of angular momentum advected inward along with the mass,

$$\text{constant} = -\dot{M} \cdot r^2 \, \Omega \simeq -\dot{M} \sqrt{G M_\star} \, R_\star^{1/2}. \tag{9.20}$$

The integration constant can thus be seen as the influx of angular momentum through the disk. With this, and $\Omega = \Omega_{\rm K} = \sqrt{GM_*/r^3}$, the disk surface density becomes

$$\Sigma = \frac{\dot{M}}{3 \, \pi \, \nu} \left( 1 - \sqrt{\frac{R_*}{r}} \right), \tag{9.21}$$
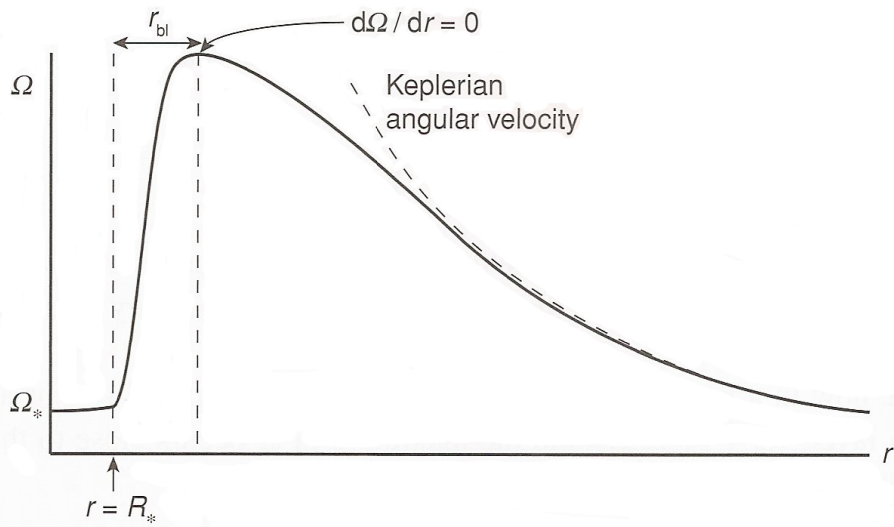
Figure 9.5: *Schematic illustration of the angular velocity profile in a disk that extends to the equator of a slowly- or non-rotating star. The viscous stress vanishes at a radial location $r = R_\star + r_{\rm bl}$, where $\Omega$ has a maximum and thus $d\Omega/dr = 0$. From: Philip J. Armitage, Astrophysics of Planet Formation.*

for radii much larger than the stellar radius. Once the viscosity is specified, this equation gives the steady-state surface density profile of a proto-planetary disk with a constant accretion rate $\dot{M}$. Away from the inner boundary one notes that $\Sigma(r) \propto \nu^{-1}$. Solutions have been derived for a power-law behavior $\nu \propto r^\gamma$ of the viscosity (see e.g. Lynden-Bell & Pringle 1974 and section 9.3).

The solution is, as pointed out, for a steady disk subject to a *zero-torque* boundary condition at the inner edge. Physically, this boundary condition is at least approximately realized for disks that extend to the equator of a slowly rotating star, and it is also the traditional choice in the more exotic circumstance of a disk of gas around a black hole. In classical T Tauri stars it is often the case that stellar magnetic fields truncate the disk before it reaches the stellar surface, and the resulting magnetic coupling between the star and the inner disk can violate the zero-torque assumption. In a disk that does not reach the stellar surface, the details of the inner disk boundary condition will determine the surface density very close to the inner edge of the disk.

### The Minimum Mass Solar Nebula model

One may attempt to constrain the radial surface density profile of the proto-solar disk from the solar system objects orbiting the sun. The classical approach is to estimate the metal contents of the planets as a starting point and to use the cosmic abundance pattern to recover the original amount of gas (so, including H and He) associated with each planet. This mass is then assumed to have been distributed in annuli centred on each planet's current location,
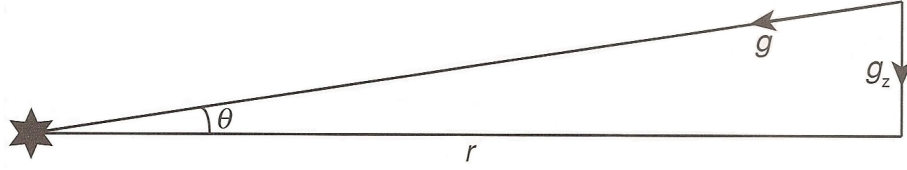
Figure 9.6: *The vertical structure of geometrically thin disks is set by a balance between the vertical component of the star's gravity $g_z$ and a pressure gradient. From: Armitage, Astrophysics of planet formation (2010). Note that the sketch is wrong in that the vertical component $g_z$ is drawn too long.*

and the surface densities are found simply by dividing the total (metals + gas) mass of each planet by the surface of an annulus the width halfway between the planet to the inside and halfway the planet to the outside. For instance, Weidenschilling (1977) corrected the mass of Jupiter ($318\,M_\oplus$) to a total mass in the range of 600 to $1.2 \times 10^4\,M_\oplus$, distributed in an annulus from 3.3 to 7.4 AU, and so estimated an initial surface density at Jupiters's location (5.2 AU) of about $120 - 2400\,\mathrm{gr\,cm^{-2}}$. Later, Hayashi (1981) repeated this calculation with differing assumptions about where different solids would condense (especially water ice) and found a minimum value of the proportionality constant consistent with the known planetary masses, deriving a commonly used Minimum Mass Solar Nebula (MMSN) surface density

$$\Sigma(r) = 1700 \frac{1}{f_p} \left( \frac{r}{1\,\mathrm{AU}} \right)^{-3/2} \qquad \mathrm{g\,cm^{-2}}, \tag{9.22}$$

with an implied range of validity from inside Mercury's orbit to beyond Neptune's orbit, about $0.3 - 30$ AU. Integration of the MMSN $\Sigma(r)$ between 0.3 and 30 AU yields a mass of $0.013\,M_\odot$ for $f_p = 1$. The factor $f_p$ accounts for the possibility that only a fraction of the solid material (that in planetesimals) was captured by planets; the remaining fraction was in the form of dust that was lost as the solar nebula cleared. Notice that this still assumes that all of the planetesimals in the disk were accreted by planets.

Of course, this density profile depends crucially on the position of the planets. A more recent model (the so-called Nice model) assumes the four giant planets to initially have been in a compact formation. Desch (2007) constructed a new MMSN, assuming the planets were formed in the disk at their starting position in the Nice model. He finds

$$\Sigma(r) = 25250 \frac{1}{f_p} \left( \frac{r}{1\,\mathrm{AU}} \right)^{-2.168}. \tag{9.23}$$

### Vertical disk structure

The structure in the direction perpendicular to the disk mid-plane, the $z$-direction, is found by solving the equation of hydrostatic equilibrium (see Fig. 9.6 and Eq. 3.23 for the static case),

$$\frac{1}{\rho} \frac{\partial p}{\partial z} = g_z = \frac{\partial}{\partial z} \left( \frac{G\,M_\star}{\sqrt{r^2 + z^2}} \right) = -\frac{GM_\star z}{(r^2 + z^2)^{3/2}}. \tag{9.24}$$

If the disk is vertically thin ($z \ll r$) and the gas in the $z$-direction at a given radius is isothermal, such that $p = a^2 \rho$ (see Eq. 3.3), equation (9.24) can be integrated to give the vertical density structure,

$$\rho(r, z) = \rho_{\mathrm{mp}}(r) \, \exp \left[ -\frac{1}{2} \left( \frac{z}{h} \right)^2 \right], \tag{9.25}$$

where $\rho(r, z = 0) = \rho_{\mathrm{mp}}(r)$ is the density at the disk mid-plane.

The *gas scale height* is

$$h = \left( \frac{kT}{\mu m_{\mathrm{p}}} \frac{r^3}{G M_\star} \right)^{1/2} = \frac{a}{\Omega}, \tag{9.26}$$

where $T$ is the gas temperature at $r$ for all $z$. The density profile is thus a Gaussian profile in the $z$-direction, peaking in the mid-plane. Furthermore, it follows that the gas scale height describing the characteristic vertical thickness is the ratio of the gas sound speed to the angular velocity. It is important to ponder a bit on the vertical density profile. Filling in some obvious numbers, we get $\rho(1h)/\rho_{\mathrm{mp}} \simeq 0.61$ and $\rho(2h)/\rho_{\mathrm{mp}} \simeq 0.14$. This explicates that the density is about constant in the first scale height above the mid-plane and that above $z = h$ the density drops quite rapidly, already be a factor of about 7 at $z = 2h$. At $z = 5h$ the disk medium is extremely thin, $\rho(5h)/\rho_{\mathrm{mp}} \simeq 3.7 \times 10^{-6}$. Using the vertical density profile Eq. (9.25) and Eq. (9.4) it follows that the mid-plane density is

$$\rho_{\mathrm{mp}} = \frac{1}{\sqrt{2\pi}} \frac{\Sigma}{h}. \tag{9.27}$$

One may introduce the *aspect ratio* $h/r \propto T_{\mathrm{mp}}^{1/2} \, r^{1/2}$, which describes the rise of the disk (or disk surface) with distance. The aspect ratio is an important parameter in describing the structure of the disk:

○ *h/r = constant.* In this case the disk has a constant opening angle. For a *thin disk*, $h/r \ll 1$. In the special case that the constant is zero one refers to the disk as a *razor-thin disk*.

○ *h/r increases with r.* The disk structure has a bowl-like shape. This is referred to as a *flaring disk*. Features of flared disks are, first, that all points on the surface of the disk have a clear line-of-sight to the star, and, second, that, seen from the star, the disk subtends a greater solid angle than a razor-thin disk.

○ *h/r decreases with r.* Though it is impossible for the disk to have negative $d(r/h)/dr$ at all radial positions, one can envision a disk that has a certain vertical height near the inner disk rim, and a smaller aspect ratio further out. If the disk is optically thick, these further out regions will be in the shadow of the inner disk rim and will not have a clear view of the star. Such a disk is referred to as being *self-shadowed*.

Let us compute the aspect ratio for a disk accreting at a low rate on to a solar mass star. For such a system Bell et al. (1977) show that the mid-plane temperature at 1 AU is around

$T_{\mathrm{mp}} \simeq 100\,\mathrm{K}$. The corresponding isothermal sound speed is $a_{\mathrm{mp}} \simeq 0.6\,\mathrm{km\,s^{-1}}$, where we have assumed a mean molecular weight $\mu = 2.3$. This yields $h/r \approx 0.02$. The condition that the disk is geometrically thin, assumed above to compute the vertical density structure, is adequately satisfied.

While checking the assumptions in our derivation, we should also verify that the vertical structure is not modified by the gravitational force from the disk itself. Representing the disk as a thin infinite sheet of mass with surface density $\Sigma$, a straightforward application of Gauss' theorem shows that the acceleration outside the sheet is constant with (vertical) distance,

$$g_{z,\mathrm{disk}} = 2\pi G\,\Sigma. \tag{9.28}$$

Equating this disk contribution to the acceleration due to the vertical component of stellar gravity at $z = h$, the condition that the self-gravity of the disk can be neglected when computing the vertical structure becomes

$$\Sigma \lesssim \frac{M_\star\,h}{2\pi\,r^3}. \tag{9.29}$$

For the minimum mass Solar Nebula at 1 AU this condition is satisfied by more than an order of magnitude. More generally, if we write the enclosed disk mass as $M_{\mathrm{disk}} = \pi\,r^2\,\Sigma$, self-gravity can be ignored provided that

$$\frac{M_{\mathrm{disk}}}{M_\star} \lesssim \frac{1}{2}\left(\frac{h}{r}\right). \tag{9.30}$$

At large radii $h/r \sim 0.1$, so *for a massive disk* with $M_{\mathrm{disk}}/M_\star \simeq 0.1$ the neglect of disk self-gravity in the calculation of the vertical structure is not justified. Such disks require additional considerations. We mention in passing that once the disk self-gravity becomes important it becomes unstable to the development of non-axisymmetric structure in the form of spiral waves.

## Radial disk structure

Having found an expression for the scale height, we return to the radial disk structure and write Eq. (9.21) using $r \gg R_\star$ as

$$\Sigma = \frac{\mu\,m_p\,\sqrt{G\,M_\star}}{3\pi\,k}\,\frac{\dot{M}}{\alpha\,T_{\mathrm{mp}}\,r^{3/2}}. \tag{9.31}$$

Here we used the relation $\nu = \alpha\,a\,h$, where $\alpha$ is a dimensionless parameter, introduced by Shakura & Sunyaev to describe viscous disks. We discuss this description of viscosity in more detail in section 9.4. The sound speed $a$ is given by Eq. (3.3), as always. For a disk with a simple power-law midplane temperature profile, $T_{\mathrm{mp}}(r) \propto r^{-q}$, the surface density is proportional to $r^{\,q-3/2}$, thus a simple power law of radius. Further, the midplane density may be written as

$$\rho_{\mathrm{mp}} \simeq \Sigma/h = \rho_{\mathrm{in}}(r/r_{\mathrm{in}})^{\frac{3}{2}q-3}, \tag{9.32}$$

where $\rho_{\mathrm{in}}$ is the mid-plane density at the inner disk radius. Assuming a simple power law for the temperature profile is a first guess, but is there actually a way to determine the temperature profile within an accretion disk?

## 9.2 Temperature profile of an accretion disk

The time scale for the disk to attain thermal equilibrium is generally much less than the time scale for evolution of either the disk or the star. The temperature profile is then set by the balance between cooling and heating, for which there are two main sources:

○ Dissipation of gravitational potential energy, as matter in the disk spirals in towards the star. Disks for which this is the main heating process are described as *active disks*.

○ Intercepted stellar radiation, which is absorbed (usually by dust) and subsequently re-radiated at longer wavelengths. Disks for which this is the main heating process are described as *passive disks*.

The heating per unit area from both of these sources drops off with increasing distance from the star. If the disk extends out to large enough radii, heating due to the ambient radiation field provided by nearby stars can become significant and prevent the disk temperature from dropping below some minimum level. For normal star-forming environments this minimum temperature might be 10–30 K.

The relative importance of accretional heating versus stellar irradiation depends upon the accretion rate and the amount of intercepted stellar radiation. Globally, in-spiralling matter will radiate an amount of energy per unit mass that is approximately given by Eq. (6.25). By comparing the accretion luminosity to the luminosity the disk receives from the central star, which for a razor-thin disk is $1/4\,L_\star$ (see below), we can derive the mass-accretion rate for which both sources contribute equally. For a young solar mass star with a luminosity $L_\star = L_\odot$ and a radius $R_\star = 2R_\odot$, one finds $\dot{M} \approx 2 \times 10^{-8}\,M_\odot\mathrm{yr}^{-1}$. Measured accretion rates for classical T Tauri stars range from an order of magnitude above this critical rate to two orders of magnitude below. So, the thermal structure of disks at early epochs (when accretion is strongest) is likely dominated by internal heating due to accretion, whereas at later times reprocessing of stellar radiation dominates.

### Temperature structure of active disks

The radial dependence of the temperature of an actively accreting disk can be derived by considering the net torque on an annulus of width $\Delta r$, as the energy generation is predominantly

through the work done by viscous torques. The net torque on a disk annulus is equal to the difference between the torque on the outer and inner surface

$$G(r + \Delta r) - G(r) = \frac{\partial G}{\partial r} \Delta r. \tag{9.33}$$

This torque does work at a rate (recall the dimension of $\Omega$ is s$^{-1}$)

$$\Omega \frac{\partial G}{\partial r} \Delta r = \left[ \frac{\partial}{\partial r} (G \, \Omega) - G \frac{d\Omega}{dr} \right] \Delta r \tag{9.34}$$

The first term on the right-hand side is simply $G\Omega|_{r_{\text{out}}} - G\Omega|_{r_{\text{in}}}$ and thus given by the disk boundary conditions. The second term on the right-hand side describes the local energy dissipation and thus the heat generation in the disk. The rate of heating by viscous processes per unit disk surface area is

$$
\begin{aligned}
Q_{\text{heating}} &= \frac{G \, d\Omega/dr}{4\pi r} \\
&= \frac{1}{2} \nu \, \Sigma \left( r \frac{d\Omega}{dr} \right)^2 \\
&= \frac{9}{8} \nu \, \Sigma \, \Omega^2 \tag{9.35}
\end{aligned}
$$

This starts out by realising that the surface area of the disk $2\pi r \, \Delta r$ needs to be multiplied by an additional factor of 2 because the disk has two sides. We have assumed that the disk has a Keplerian rotation profile $\Omega = \sqrt{GM/r^3}$, such that $d\Omega/dr = -3\Omega/2r$. Using the relation found between the mass accretion rate and the viscosity, $\dot{M} = 3\pi\nu\Sigma$ for $r \gg R_*$, we can eliminate the viscosity from this equation

$$Q_{\text{heating}} = \frac{3}{8\pi} \dot{M}\Omega^2. \tag{9.36}$$

If the disk emits as a black body, we take $Q_{\text{heating}} = Q_{\text{radiation}} = 2\sigma T_{\text{mp}}^4$ with $\sigma$ being the Stefan-Boltmann constant. The factor two enters because the disk has two sides. This yields a temperature profile of

$$T_{\text{mp}} = \left( \frac{3}{16\pi\sigma} \dot{M}\Omega^2 \right)^{1/4}, \tag{9.37}$$

thus a $r^{-3/4}$ profile. We see that the disk temperature does not depend on the viscosity. The implicit assumption here is that the viscosity of the disk corresponds to its accretion rate, an observable quantity. Hence we do not need to know the details of the angular momentum transport to describe the disk structure. Using a typical observed accretion rate for young T Tauri stars ($M_\star = 1\ M_\odot$) of $2 \times 10^{-7}\ M_\odot\text{yr}^{-1}$, we obtain a disk temperature of 150 K at a distance of 1 AU.
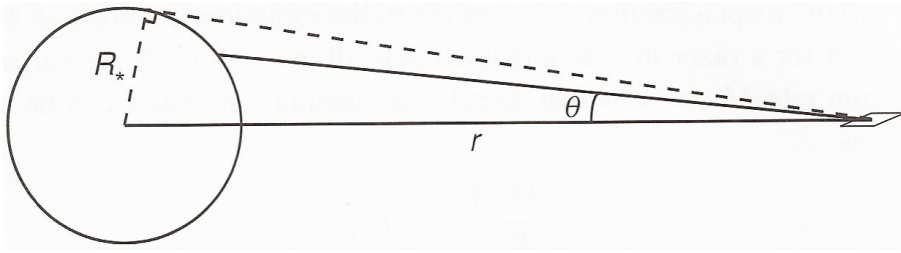
Figure 9.7: *Geometry for the calculation of the radial temperature profile of a razor-thin proto-planetary disk. The flux impinging on the surface of the disk at distance r from the star is computed by integrating over the visible stellar surface. From: Armitage, Astrophysics of planet formation (2010).*

## Temperature structure of razor-thin passive disks

The temperature profile of a passive disk is determined by the shape of the disk, i.e. whether it is flat, flared, or warped, and by the mechanism by which the absorbed stellar radiation is re-emitted. Let us first consider the simplest model: a razor-thin disk that absorbs all incident radiation and re-emits it locally as a single temperature black body. Back-warming of the star by the disk is neglected.

We consider a surface in the plane of the disk at distance $r$ from a star of radius $R_\star$. The star is assumed to be a sphere of constant total surface brightness $I_\star$. We adopt a spherical polar coordinate system such that the axis of the coordinate system points to the center of the star, as shown in Fig. 9.7, the stellar flux passing through this surface is

$$\mathcal{F} = \int I_\star \sin\theta \, \cos\phi \, d\Omega, \tag{9.38}$$

where $d\Omega$ represents the element of solid angle. The sine and cosine terms account for projection effects. If one is located on the stellar surface, in the line-of-sight of direction $(\theta, \phi)$ as seen from the small disk surface, one needs to compute the projection of that surface to determine how much light is actually absorbed. If one is in $(0,0)$ the term $\sin\theta = 0$ implies that – as seen from the stellar surface – the small surface in the plane of the disk is viewed on its side (so it has no projected surface). For a non-zero $\theta$, the term $\cos\theta$ too corrects for the proper projection. For both a 'rolling angle' $\phi = \pi/2$ and $-\pi/2$ again the disk surface is viewed on its side. We count the flux coming from the top half of the star only – and to be consistent equate that to radiation from only the top surface of the disk. So, integrating over the stellar surface, the limits of the integral are $-\pi/2 < \phi < +\pi/2$ and $0 < \theta < \sin^{-1}(R_\star/r)$. Substituting $d\Omega = \sin\theta \, d\theta \, d\phi$, the integral for the total flux is

$$\mathcal{F} = I_\star \int_{-\pi/2}^{+\pi/2} \cos\phi \, d\phi \int_0^{\sin^{-1}(R_\star/r)} \sin^2\theta \, d\theta. \tag{9.39}$$

For a star with effective temperature $T_\star$, the total surface brightness $I_\star = \sigma T_\star^4/\pi$. Equating

$\mathcal{F}$ to the one-sided disk emission $\sigma T^4$ we obtain a radial temperature profile

$$\left(\frac{T}{T_\star}\right)^4 = \frac{1}{\pi}\left[\sin^{-1}\left(\frac{R_\star}{r}\right) - \left(\frac{R_\star}{r}\right)\sqrt{1 - \left(\frac{R_\star}{r}\right)^2}\right]. \tag{9.40}$$

Integrating over radii, we obtain the total disk luminosity

$$L = 2\int_{R_\star}^{\infty} 2\pi\, r\, \sigma T^4\, dr = \frac{1}{4}L_\star. \tag{9.41}$$

So, a flat passive disk extending all the way to the stellar equator intercepts a quarter of the stellar flux. Taking the viewpoint of a distant observer of the star + disk system one may wonder what is the ratio of the observed luminosity of such a disk to the stellar luminosity. This will vary with viewing angle, but clearly a flat passive disk is predicted to be less luminous than the star.

The exact behavior of the temperature profile given by Eq. (9.40) is not very transparent. Expanding the right-hand side in a Taylor series in the limit that $R_\star/r \ll 1$ (i.e. far from the stellar surface), we obtain

$$T \propto r^{-3/4}, \tag{9.42}$$

as the limiting temperature profile of a razor-thin passive disk. Assuming that in the vertical direction the disk is isothermal, the aspect ratio $h/r \propto r^{1/8}$, and we thus predict that the disk ought to flare modestly to larger radii. If the disk does flare then the outer regions intercept a larger fraction of stellar photons, leading to a higher temperture. As a consequence, a temperature profile $T \propto r^{-3/4}$ is probably the steepest profile we would expect to obtain for a passive disk.

## Temperature structure of flared passive disks

The temperature structure of a passive flared disk can be computed in the same way as for a passive razor-thin disk, namely by evaluating the flux Eq. (9.38) by integrating over the stellar surface accounting for the projection of the disk surface at radius $r$. The exact calculation can be found in the appendix to Kenyon & Hartmann (1987), and while it is conceptually simple some messy geometry is required. Here, we adopt an approximate treatment valid for $r \gg R_\star$, and consider the star to be a point source of radiation. At radial distance $r$ in the mid-plane of the disk, stellar radiation is absorbed by the disk at height $h_{\mathrm{p}}$ above the mid-plane. Note that $h_{\mathrm{p}}$ is not the same as the disk scale-height $h$ as the location of the disk surface is determined by the (radial) optical depth of the gas and not by the (vertical) density structure. From considerations of geometry in Fig. 9.8, the angle between the incident radiation and the local disk surface is given by the difference between the angle representing the radial slope of the actual disk surface minus the angle representing the slope of the incident radiation hitting the disk surface at height $h_{\mathrm{p}}(r)$, i.e.

$$\alpha = \frac{\mathrm{d}h_{\mathrm{p}}}{\mathrm{d}r} - \frac{h_{\mathrm{p}}}{r}. \tag{9.43}$$
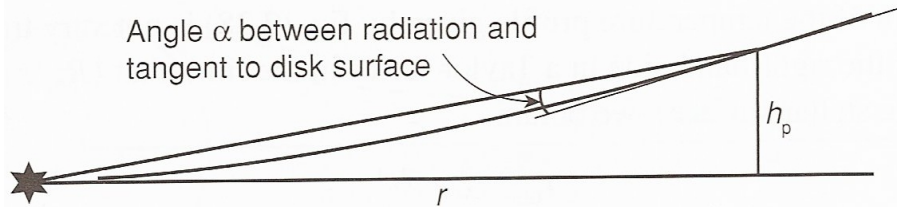
Figure 9.8: *Geometry for calculation of the radial temperature profile of a flared proto-planetary disk. At distance $r \gg R_\star$, radiation from the star is absorbed by the disk at height $h_\mathrm{p}$ above the mid-plane. The angle between he tangent to the disk surface and the incident radiation is $\alpha$. From: P.J. Armitage, Astrophysics of planet formation.*

The rate of heating per unit disk area at distance $r$ is

$$Q_\mathrm{heating} = 2\alpha \left( \frac{L_\star}{4\pi \, r^2} \right), \tag{9.44}$$

where the factor of two comes from the fact that the disk has two sides and we have assumed that all of the stellar surface is visible from the surface of the disk and that $\alpha$ is a small angle. Equating the heating rate to the rate of cooling by blackbody radiation $Q_\mathrm{cooling} = 2\sigma T^4$, yields for the disk temperature profile

$$T = \left( \frac{L_\star}{4\pi\sigma} \right)^{1/4} \alpha^{1/4} \, r^{-1/2}. \tag{9.45}$$

Since $L_\star = 4\pi R_\star^2 \, \sigma T_\star^4$ an equivalent expression is

$$\frac{T}{T_\star} = \left( \frac{R_\star}{r} \right)^{1/2} \alpha^{1/4}. \tag{9.46}$$

If we assume that $h_\mathrm{p} \propto h$, which may be plausible if the disk is very optically thick – then Eq. (9.26), (3.3), (9.43), and (9.46) can be solved to yield $\alpha$. At large radii, Kenyon & Hartmann (1987) find that the surface temperature approaches

$$T(r) \propto r^{-1/2}, \tag{9.47}$$

which is, as expected, flatter than the $T \propto r^{-3/4}$ profile of a flat disk (Eq. 9.42).

## 9.3  Angular momentum transport

The specific angular momentum stored in a 1 $M_\odot$ disk with a size of 10 AU is $3 \times 10^{53}$ cm$^2$/sec. A 1 $M_\odot$ star rotating at break-up velocity has a specific angular momentum of $6 \times 10^{51}$ cm$^2$/sec.

Hence, the original angular momentum of the disk is 50 times higher than the maximum allowed for a star. Since angular momentum is strictly conserved, there needs to be a process that transports angular momentum away to prevent it from accumulating on the star. The main possibilities to do this are

- ○ a torque from the external medium (e.g. magnetic fields),

- ○ viscosity inside the disk transporting angular momentum to the outer disk,

- ○ disk winds taking angular momentum away.

In the following, we discuss the basic principle of angular momentum transport within the disk. In a Keplerian disk (see Fig. 9.9) friction between a parcel of material in ring $A$ and a parcel of material in ring $B$ will try to slow down parcel A and speed up parcel $B$, as $A$ is moving faster than $B$. If parcel $A$ looses angular momentum, but is forced to remain on a Kepler orbit, it must move inward. Parcel $B$ moves outward, unless it, too, has friction with a parcel $C$ (in a ring $C$ outside of $B$), which in turn may have friction with a parcel $D$ (in a ring $D$ outside of $C$), etcetera. In this way, momentum is transported out through the disk, while material



Figure 9.9: *Schematic view of velocities in a Keplerian disk.*

is moving in. In a more mathematical way we may describe this by considering the angular momentum

$$
\begin{aligned}
J &= m_A \, v_{\phi,A} \, r_A + m_B \, v_{\phi,B} \, r_B \\
&= \sqrt{GM_*} \, (m_A \, \sqrt{r_A} + m_B \, \sqrt{r_B}),
\end{aligned}
\tag{9.48}
$$

where $r_A$ and $r_B$ are the corresponding distances of the two masses from the central star with mass $M_\star$. From the conservation of angular momentum of the total system, it follows that a small change in orbit for one of the masses requires a corresponding change for the other one

$$
\begin{aligned}
\frac{\partial J_A}{\partial r} &= -\sqrt{GM_*} \, m_A \frac{1}{2\sqrt{r_A}} \quad \rightarrow \quad \Delta J_A = -\sqrt{GM_*} \, m_A \frac{\Delta r_A}{2\sqrt{r_A}} \\
m_A \frac{\Delta r_A}{\sqrt{r_A}} &= -m_B \frac{\Delta r_B}{\sqrt{r_B}}
\end{aligned}
\tag{9.49}
$$

The same of course holds true if we consider two neighboring annuli in the disk.

### Analytical solutions of mass and angular momentum transport

Let us investigate the essential concept of angular momentum transport by inspecting a time-dependent analytic solution to Eq.(9.15) adopting a simple prescription for the viscosity, even though this prescription is not particularly realistic for proto-planetary disks.

Figure 9.10: *The Green's function solution to the disk evolution equation with $\nu$ = constant, showing the spreading of a ring of mass initially orbiting at $r = r_\circ$. From top down the curves show the behavior as a fucntion of the scaled time variable $\tau$ = 0,004, 0.008, 0.016, 0.032, 0.064, 0.128, and 0.256. From: P.J. Armitage, Astrophysics of planet formation.*
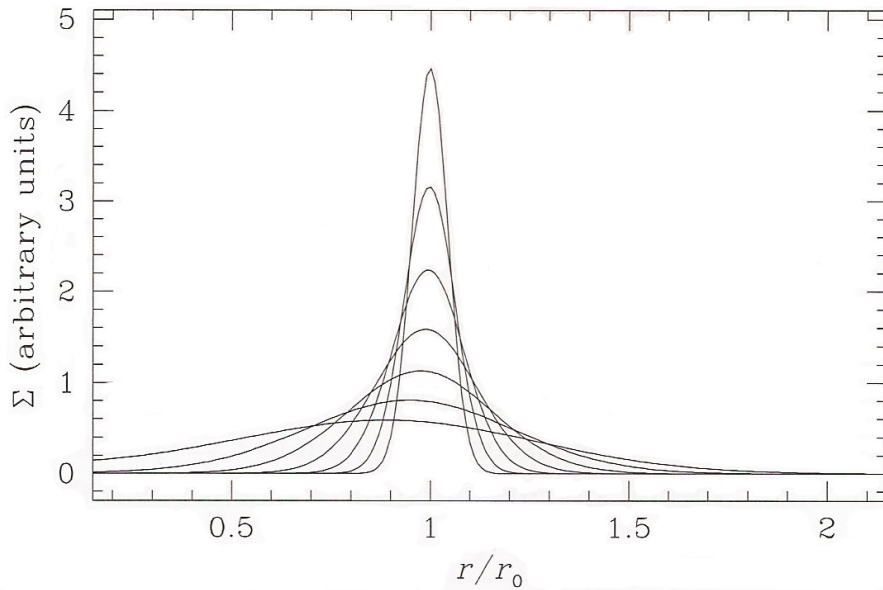
If $\nu$ = constant, a Green's function solution to the evolution equation is possible. We assume that at $t = 0$, all of the gas lies in a thin ring of mass $m$ at radius $r_\circ$. This means that

$$\Sigma(r, t = 0) = \frac{m}{2\pi\, r_\circ}\, \delta\left(r - r_\circ\right), \tag{9.50}$$

where $\delta(r - r_\circ)$ is a Dirac delta function. Given boundary conditions that impose zero-torque at $r = 0$ and allow for free expansion toward $r = \infty$ the solution is (see Lynden-Bell & Pringle, 1974 for the details)

$$\Sigma(x, \tau) = \frac{m}{\pi\, r_\circ^2}\, \frac{1}{\tau}\, x^{-1/4} \exp\left[-\frac{(1 + x^2)}{\tau}\right] I_{1/4}\left(\frac{2x}{\tau}\right), \tag{9.51}$$

where $x \equiv r/r_\circ$ and $\tau = 12\,\nu\, r_\circ^{-2}\, t$. Both these variables are dimensionless. The quantity $I_{1/4}$ is a modified Bessel function of the first kind. The solution is also plotted in Fig. 9.10 and shows several generic features of disk evolution. As $t$ increases the initially narrow ring spreads diffusively, with the mass flowing toward $r = 0$ while simultaneously the angular momentum is carried by a small fraction of the mass toward $r = \infty$. This segregation of mass and angular momentum is called *viscous spreading of the disk* and is a general feature of the evolution of a viscous disk, and is necessary if accretion is to proceed without overall angular momentum loss from the system.

Of greater practical utility than the Green's function solution is the self-similar solution also derived by Lynden-Bell & Pringle (1974). These authors consider a disk in which the viscosity
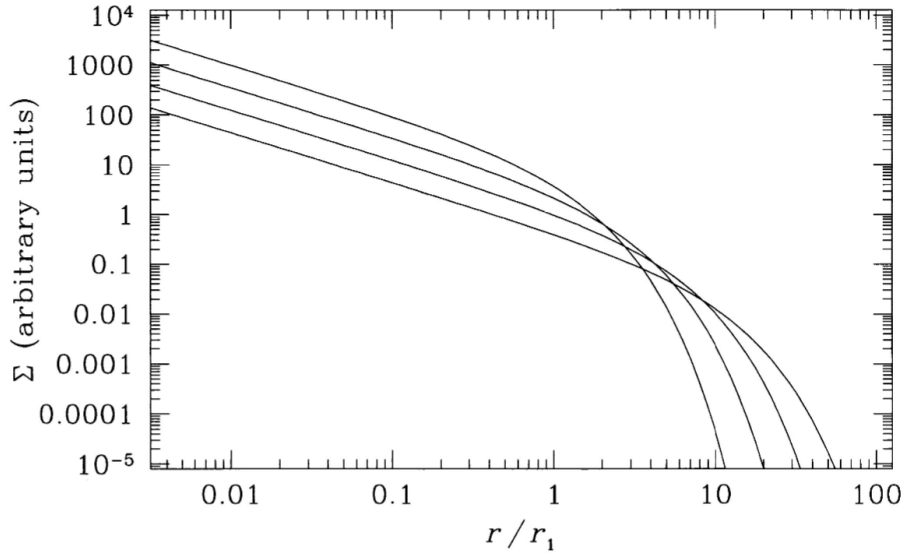
Figure 9.11: *The self-similar solution to the disk evolution equation is plotted for a viscosity $\nu \propto r$. The initial surface density tracks the profile for a steady-state disk ($\Sigma \propto r^{-1}$), before cutting off exponentially beyond $r = r_1$. The curves show the surface density at the initial value of the scaled time $\tau = 1$ (see definition equation 9.54), and at subsequent times $\tau = 2, 4$ and 8. From: P.J. Armitage, Astrophysics of planet formation.*

can be approximated as a power-law in radius,

$$\nu \propto r^{\gamma}. \tag{9.52}$$

Applying the Shakura & Sunyaev viscosity prescription $\nu = \alpha\, h\, r$ one can work out that $\gamma = 3/4$ for the temperature profile $T \propto r^{-3/4}$ (see Eq. 9.37 or 9.42). For the temperature structure $T \propto r^{-1/2}$, more typical for a passive flaring disks (see Eq. 9.47), one obtains $\nu \propto r$. Let us, for the remainder of the discussion, adopt this latter description, that is, we use $\gamma = 1$.

Suppose that the disk at time $t = 0$ has the surface density profile corresponding to steady-state solution Eq. (9.21) at $r \gg R_{\star}$ out to $r = r_1$, with an exponential cut-off at larger radii. Specifically, the initial surface density has the form

$$\Sigma(t = 0) = \frac{C}{3\pi\, \nu_1\, (r/r_1)} \exp\left[-r/r_1\right], \tag{9.53}$$

where $C$ is a normalization constant and $\nu_1 = \nu(r_1)$. The self-similar solution is then

$$\Sigma(r, \tau) = \frac{C}{3\pi\, \nu_1\, (r/r_1)}\, \tau^{-3/2} \exp\left[-\frac{(r/r_1)}{\tau}\right], \tag{9.54}$$

where $\tau = (t/t_{\rm s}) + 1$ is the scaled time and $t_{\rm s} = (1/3)\,(r_1^2/\nu_1)$. The evolution of related quantities such as the disk mass and accretion rate can readily be derived from the above
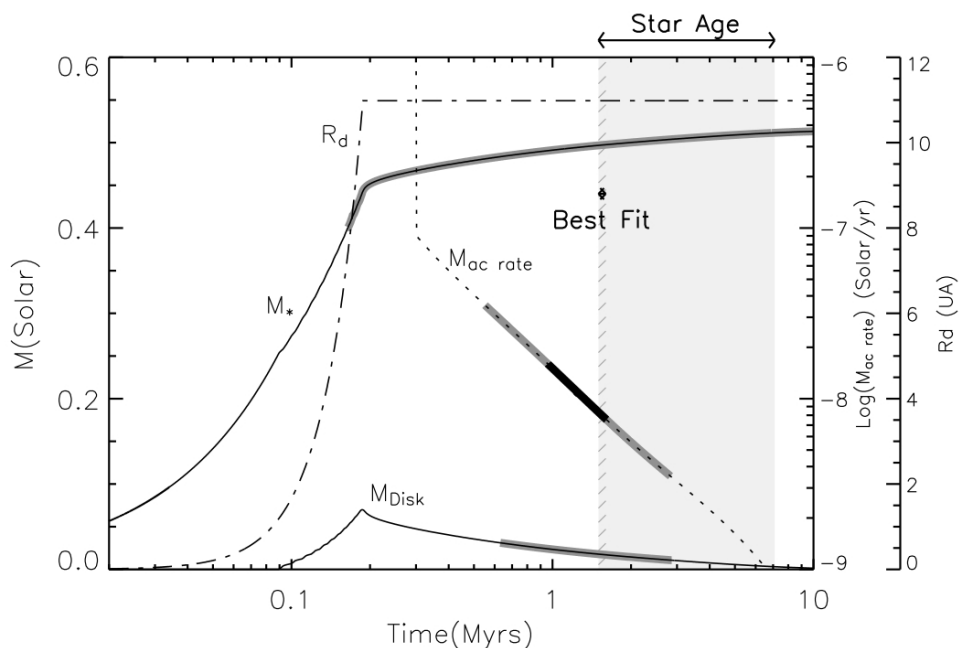
Figure 9.12: *Evolution of star mass $M_\star$ and disk mass $M_{\text{disk}}$ as a function of time with masses in solar units (corresponding axis to the left) for parameters representative of DM Tau and an alpha parameter $\alpha = 0.01$. The accretion rate onto the central star is shown as a dotted line (corresponding axis: first to the right). The evolution of the centrifugal radius $r_{\text{cf}}$ (in the figure labelled $R_{\text{d}}$) is shown as a dash-dotted line (corresponding axis: far-right; note that UA should have been AU). Gray and black stripes indicate properties consistent with the current values of the DM Tau system, of which the likely age is indicated with the grey zone. All constraints agree within the errors for a system age of 1.5–3 Myr. From: Hueso & Guillot, 2005, A&A 442, 703.*

expression for the surface density. The solution is plotted in Fig. 9.11. Over time, the disk mass decreases because of accretion onto the star while the characteristic scale of the disk (initially $r_1$) expands to conserve angular momentum.

## Numerical solution of mass and angular momentum transport

Hueso & Guillot (2005) present a one-dimensional model of the formation and viscous spreading of proto-planetary disks. One parameterization of the viscosity applied in this study is that of the classical $\alpha$ representation by Shakura & Sunyaev. They apply their model to DM Tau and GM Aur, two classical T-Tauri stars with relatively well-characterized disks. Using empirical constraints on stellar and disk properties they constrain the value of $\alpha$ and investigate whether viscous spreading can actually 'grow' disks from initially small centrifugal radii of several AU (see Eq. 7.8) to the observed radii of 850 AU for DM Tau and 525 AU for GM Aur in the proper evolutionary time scale (of 1–10 Myr). For instance, they find that values of $\alpha$ in

the range $0.001 - 0.1$ are compatible with the observed disk surface density distribution, star and disk mass, age and present accretion rate.

Figure 9.12 shows their result for DM Tau. The thick grey and black stripes overlaying the different quantities denote the best observational values (with black being more certain than grey). This example shows how DM Tau's 850 AU disk can be formed by viscous diffusion of an initially much smaller disk, with a centrifugal radius $r_{cf} = 11$ AU. At the end of the simulation, the star has almost acquired its final mass (here 0.5 $M_\odot$ in 10 Myr). The disk mass grows to about 0.1 $M_\odot$ in $\sim 2 \times 10^5$ yr. At $\sim 3 \times 10^5$ yr the initial cloud has largely dispersed and the mass-accretion rate onto the star drops sharply to less than $10^{-7}$ $M_\odot \mathrm{yr}^{-1}$. The disk mass decreases in time as its material is slowly accreted onto the star.

## 9.4  The nature of viscosity

We have so far postponed a discussion of the physical origin of angular momentum transport. So, lets ask ourselves the relevant question: what is the nature of viscosity?

### Viscous timescale

Before doing so, we introduce the *viscous timescale* $\tau_{vis}$ using a dimensional argument. As the unit of viscosity is $\mathrm{cm}^2 \sec^{-1}$, one may anticipate that the time scale on which viscosity will smooth out surface gradients on a spatial scale $\Delta r$ is $\tau \sim (\Delta r)^2/\nu$. If the disk has a characteristic size $R_{out}$, the surface density at all radii will evolve on a time scale $R_{out}^2/\nu$. Out to a distance $r$ in the disk, we could characterize the viscous timescale as

$$\tau_{vis} \approx \frac{r^2}{\nu}. \tag{9.55}$$

A more formal derivation can be performed rewriting the surface density evolution equation (9.15) into a diffusive form by introducing $x = 2r^{1/2}$ and $f = (3/2)\,\Sigma\,x$, and assuming that the viscosity is a constant. In that case the evolution equation reduces to the proto-typical form

$$\frac{\partial f}{\partial t} = D\,\frac{\partial^2 f}{\partial x^2}, \tag{9.56}$$

with $D = 12\,\nu/x^2$. The dimension of $D$ is $\sec^{-1}$, therefore the associated timescale to this equation is $\tau = 1/D = x^2/12\,\nu = (1/3)\,r^2/\nu$, similar to Eq. (9.55).

The viscous timescale can be estimated observationally by measuring, for example, the rate at which accretion on to the star decays as a function of stellar age. For proto-planetary disks around solar-type stars it appears to be a few million years as the lifetime of disks is $3 - 10$ million years (see Fig. 9.13).

$$f_{disk} = \exp(-t/\tau_{disk})$$
$$\tau_{disk} = 2.5 \text{ Myr}$$
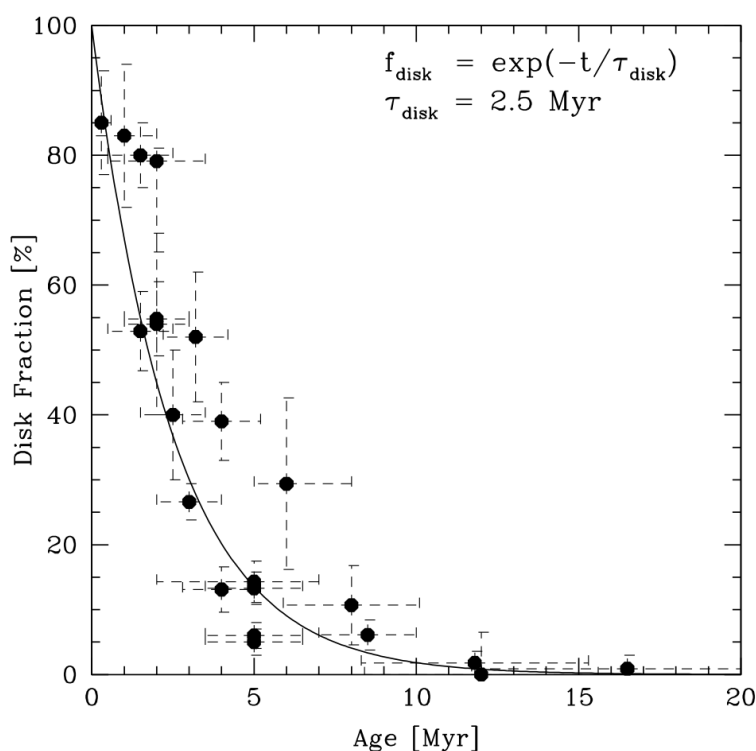
Figure 9.13: *Fraction of disk-bearing stars in a cluster as a function of average age from the birthline. The proto-planetary disk fraction for each cluster is determined by measuring either the presence of a near-infrared excess (at 3.6 μm), indicative of dust close to the star, or the presence of Hα line emission, an indicator of gas close to the star. This figure illustrates two important points. First, most stars (∼80 percent) are born with disks, and the typical lifetime of a proto-planetary disk is several (let's say 3 − 10) million years. From: Mamajek 2009.*

## Viscosity connected to turbulence: the Shakura–Sunyaev $\alpha$ parameter

The first candidate to consider is molecular collisions, which generate a viscosity in a shear flow because of the finite mean-free path in the gas. Molecular viscosity, which is the normal mechanism considered in terrestrial fluids modeled – think of syrup or honey – is totally ineffective in proto-planetary disks, as the associated timescale is on the order of $10^{13}$ yr. This is approximately ten million times longer than the observed timescale for disk evolution.

If only molecular viscosity would be present, adjacent layers would 'slide' past one another like playing cards and no cross currents perpendicular to the direction of the flow, for instance eddies or swirls, would occur. In other words, the flow would remain laminar. If, however, some form of instability is present the flow could easily produce turbulence, creating such perpendicular motion, resulting in the macroscopic mixing of fluid elements at neighbouring radii. Here it is written 'easily' as the lack of molecular coupling implies that the flow does not have any 'resistance' to changes in motion (so a lack of inertial forces gives room to

turbulence induced by viscous forces).

This turbulent viscosity can be parameterised by linking it to the maximum size of turbulent eddies and the maximum velocity of turbulent swirling motion. If the turbulence is approximately isotropic, the scale of the turbulent flow is limited to be no longer than the smallest typical scale in the disk, which is generally the disk scale-height $h$. The velocity of the turbulent motions can be similarly limited to be no larger than the sound speed $a$, since supersonic motions result in shocks and rapid dissipation (this is true even if the fluid is magnetized). We therefore write the turbulent viscosity in the form

$$\nu = \alpha \, a \, h = \alpha \, \frac{a^2}{\Omega}, \tag{9.57}$$

where $\alpha$ is a dimensionless quantity, known as the Shakura–Sunyaev $\alpha$ parameter, that measures the efficiency of angular momentum transport due to turbulence (Shakura–Sunyaev 1973). Given the way this description is designed, it must hold that $\alpha < 1$. The latter equality follows from Eq. (9.26).

It is useful to first note some of the limitations of any disk model constructed using the $\alpha$ prescription. First, the physical argument that we gave for Eq. (9.57) reasonably limits $\alpha$ to be less than unity, but it does *not* give any basis for assuming that $\alpha$ is a constant. We take $\alpha$ to be constant only as a means to proceed along lines of simple argumentation, not for any deeper reason. In fact, $\alpha$ may vary with temperature, density, and composition of the disk gas, and there may even be regions which fail to satisfy the basic assumption by not developing turbulence at all. Second, although the turbulent viscosity has the same dimensions as a molecular viscosity, one should not forget that it arises from an entirely distinct physical process.

Combining Eqs. (9.55), (9.57) and (9.26) yields

$$\tau_{\text{vis}} = \frac{r^2}{\nu} = \left(\frac{h}{r}\right)^{-2} \frac{1}{\alpha \, \Omega}. \tag{9.58}$$

If we fill in 1 Myr as the typical evolutionary timescale at 50 AU and assume that the disks are indeed very thin and $h/r \sim 0.05$, we find $\alpha = 0.02$.

So, what type of instability drives the turbulence?

### The magneto-rotational instability

Weak magnetic fields in the gas can violently de-stabilize the disk. In the presence of magnetic fields, the field lines act like springs connecting different annuli within the disk. If a weak pull exists between two masses in adjacent annuli of the disk, the inner mass element will loose angular momentum and hence drift even further inward, while the outer mass element gains angular momentum and drifts further outward. The new orbital velocity of the inner mass element is even higher than before (presuming that $\Omega$ decreases with $r$), while the new orbital

velocity of the outer element is even smaller. This effect thus enhances the velocity difference between the two masses in adjacent annuli, giving rise to an instability. The instability is generally referred to as *magneto-rotational instability* (MRI) and has been described first by Balbus & Hawley (hence it is often referred to as *Balbus-Hawley instability*). A prerequisite for a magnetized disk to be linearly unstable is that the orbital velocity decreases with radius

$$\frac{d\Omega}{dr} < 0, \tag{9.59}$$

a condition satisfied in Keplerian disks. In addition, the disk needs to be ionised to a certain degree since neutral gas does not couple efficiently to the magnetic field lines. The critical ionisation degree to sustain turbulence generation is $n_e/n_{tot} \sim 10^{-12}$ (Sano & Stone 2002). We get back to the last point in the section on 'dead zones'.

It is very difficult to assess the amount of viscosity generated by MRI. The only way of measuring $\alpha$ would be through numerical magneto-hydrodynamical simulations. Disentangling physical and numerical effects then becomes a challenge. So far simulations seem to indicate that $\alpha$ could be consistent with a value of $\sim 10^{-2}$, as inferred from observations.

## Dead zones

What might happen when the degree of ionization drops below the critical value $n_e/n_{tot} \sim 10^{-12}$ for the magneto-rotational instability to operate? Although the critical value is very small it is by no means obvious that even this level of ionization is attained throughout the disk.

Potential sources of ionization are thermal ionization near the inner rim of the disk that is exposed to direct stellar radiation, and X-ray and cosmic ray ionization. Thermal ionization of alkali metals would yield sufficient ionization for temperatures $T \gtrsim 10^3$ K. Temperatures this high are only attained in the very inner disk. In most cases, the low level of ionization arising from cosmic rays or X-rays deep inside the disk is sufficient to sustain turbulence through MRI. Cosmic rays can penetrate column densities of $\sim 50\,\mathrm{g\,cm}^{-2}$ and thus reach the mid-plane except for the inner few AU in a very massive disk with a steep surface density profile. X-rays penetrate column densities of $\sim 100\,\mathrm{g\,cm}^{-2}$. However, stellar X-rays – coming from the corona – will penetrate the disk radially and quickly reach these column densities. Only X-rays coming from high above the disk (e.g. from a stellar jet or from the accretion flow onto the star) can be expected to reach the mid-plane also at larger radii.

We conclude that close to the mid-plane the MRI at the radii of greatest interest for planet formation may be suppressed by the low ionization fraction. Unless there are other non-magnetic drivers of turbulence the disk may be quiescent with a low (or vanishing) efficiency of angular momentum transport. Regions in which the viscosity drops strongly are referred to as *dead zones*. Disk annuli experiencing MRI have lower surface densities than their quiescent (neutral and laminar) neighbors if the overal flow is in steady state, since the MRI-active regions' stronger accretion stresses drive the gas through faster. Thus, the MRI-dead zone
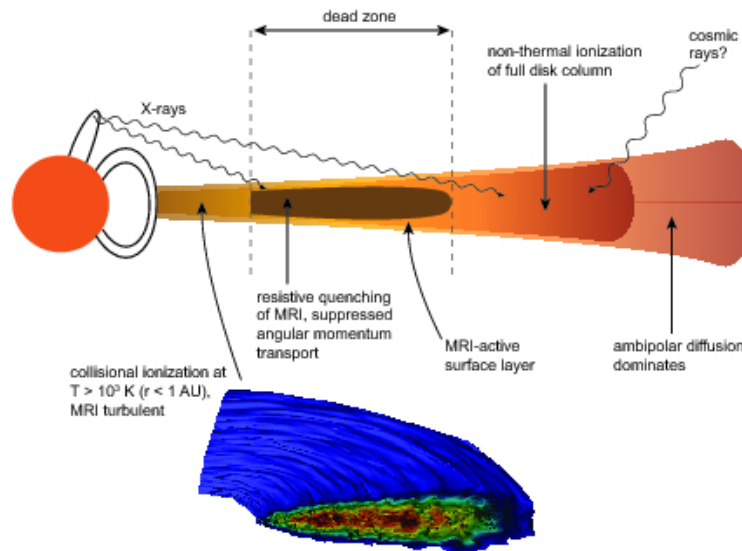
Figure 9.14: *Illustration of the layered disk model proposed by Gammie (1996). The innermost regions of the disk are hot enough such that thermal ionization suffices to couple the magnetic field to the gas well enough for the MRI to operate. At larger radii, cosmic rays and/or stellar X-rays penetrate deep enough into the disk to cause the necessary level of non-thermal ionisation. At intermediate radii, accretion occurs primarily in an active surface layer ionised by these non-thermal processes, while the central 'dead zone' is magnetically inactive. From: Phil Armitage 2010. The inset shows a global simulation of a magnetically active disk from Beckwith, Armitage, and Simon (2011).*

interface closest to the star leads to a nearby local maximum in the gas pressure (e.g. Flock et al. 2016). A gas-pressure maximum implies a pressure gradient inversion that can lead to the accumulation of small dust particles, increasing locally the dust-to-gas mass ratio in the disk (see also Sect. 10.2).

This line of reasoning leads to the *layered disk* model proposed by Gammie (1996), which is shown schematically in Fig. 9.14. In the original version of the model, the proto-planetary disk at intermediate radii (from a few times 0.1 AU out to of the order of 10 AU) has a sandwhich structure. Accretion occurs only via an active surface layer which is well enough ionized by cosmic rays or stellar X-rays to sustain MHD turbulence driven by the MRI. The column density of this active layer is of the order of $10^2 \, \mathrm{g \, cm^{-2}}$. Closer to the mid-plane lies a quiescent dead zone within which the MRI is suppressed due to insufficient ioniziation. It is assumed that there are no other sources of turbulence or angular momentum transport in the dead zone, so this region is inactive and does not support accretion. Material flows in from the outer disk and accumulates in the dead zone.

## 9.5 Gravitational instabilities

For massive disks, self-gravity can become important and introduce another sort of instability. In general, self-gravity will tend to locally clump material together. This process will be counteracted by local gas pressure or by shear motions. Gravity wins this competition if the 'clumping' timescale is short compared to the timescales on which sound waves cross a clump or shear can destroy it. The three timescales can be written for a clump of size $\Delta r$ and mass $m \sim \pi(\Delta r)^2 \Sigma$ as

$$\tau_{\text{ff}} \quad \sim \quad \sqrt{\frac{(\Delta r)^3}{Gm}} \sim \sqrt{\frac{\Delta r}{\pi G \Sigma}} \qquad (9.60)$$

$$\tau_p \quad = \quad \frac{\Delta r}{a} \qquad (9.61)$$

$$\tau_{\text{shear}} \quad = \quad \frac{1}{r}\left(\frac{d\Omega}{dr}\right)^{-1} \sim \Omega^{-1} \qquad (9.62)$$

The first equation results from a dimensional analysis of the free-fall time Eq. (3.78); the second equation from the timescale for a sound wave to cross the clump; and the third equation states the timescale required for a clump to be sheared azimuthally by an amount $\Delta r$. It is derived from the timescale associated with shear motion $\tau_{\text{shear}} = 1/\{dv_\phi/dr\}_{\text{shear}}$ (see Eq. 9.9). Note that the shear timescale is independent of the clump size $\Delta r$.

Pressure tends to stabilize small regions of the disk against gravitational collapse, while shear stabilizes the largest scales. The disk will be marginally unstable to clump formation if the free-fall timescale on the scale where shear and pressure support match is shorter than either $\tau_p$ or $\tau_{\text{shear}}$. So, eliminating the clump size using $\Delta r = a/\Omega$ and requiring $\tau_{\text{ff}} < \tau_{\text{shear}}$, we obtain for the condition for gravitational instability to occur

$$Q = \frac{a\,\Omega}{\pi G\,\Sigma} < 1. \qquad (9.63)$$

The parameter $Q$ is referred to as the *Toomre Q parameter* (Toomre 1964).

Using that $h = a/\Omega$ (Eq. 9.26) and estimating the disk mass as $M_{\text{disk}} \sim \pi\,r^2\,\Sigma$, we can write the instability threshold in the more intuitive form

$$\frac{M_{\text{disk}}}{M_\star} \gtrsim \frac{h}{r}. \qquad (9.64)$$

This is (up to numerical factors) the same condition that we derived earlier (see Eq. 9.30) for when disk gravity becomes important for the vertical structure. Since $h/r \approx 0.05$ is a representative number for proto-planetary disks, we require fairly massive disks before the effects of self-gravity can be expected to become important. Such disks are more likely to have existed at early epochs, possibly prior to the optically visible T Tauri phase of YSO evolution.

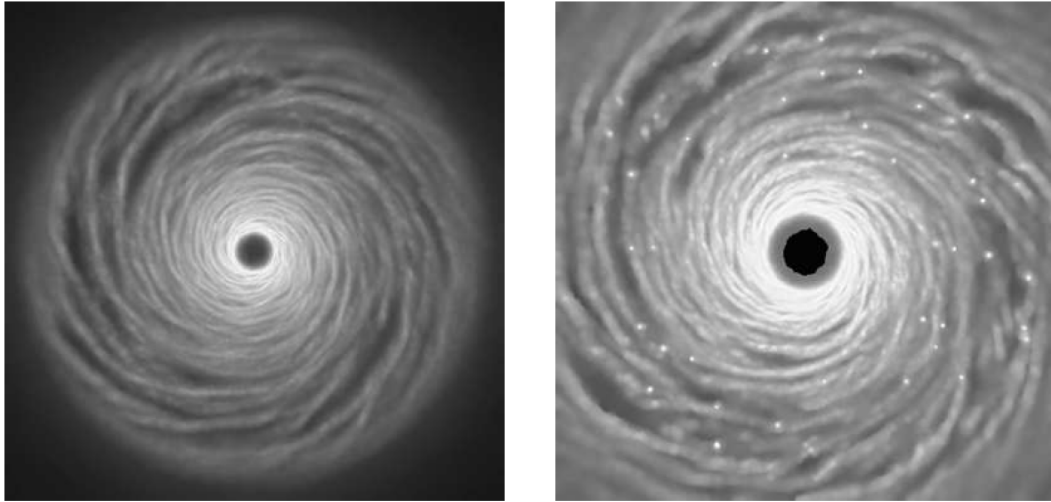If self-gravity sets in within a disk, there are two possible outcomes:

Figure 9.15: *Equatorial density structure $\rho_{\mathrm{mp}}$ for a disk of mass $M_{\mathrm{disk}} = 0.1\,M_\star$. The calculations are essentially scale-free. If we adopt a mass scale of 1 $M_\odot$ and a length scale of 1 AU, the central star would have a mass of 1 $M_\odot$, the circumstellar disc would have a mass of 0.1 $M_\odot$, and would extend from 1 AU to 25 AU. The size of the left image would than be $50 \times 50$ AU; that of the right picture $16 \times 16$ AU. Left panel: $\tau_{\mathrm{cool}} = 5\,\Omega^{-1}$. The disk is highly structured with the gravitational instability existing at all radii. The density has, however, not increased significantly and the disk is in a quasi-stable state with heating through viscous dissipation balancing cooling. Right panel: $\tau_{\mathrm{cool}} = 3\,\Omega^{-1}$. The disk is highly unstable and fragmenting. The fragments are all gravitationally bound. In both panels the gravitational instability grows from the inner radii and moves to larger radii with time. From: Rice et al. 2003, MNRAS 339, 1025.*

○ Collapse may continue unhindered, destroying the disk and forming one or more bound objects. Disk fragmentation via this process is a mechanism for giant planet, brown dwarf or binary formation, and, planetesimal formation in the case where the collapse occurs in the solid component of the disk material.

○ If collapse is hindered because the disk (including the collapsing fragments) can not radiate away its thermal energy, adiabatic heating of the clump may yield enough pressure to prevent complete collapse. Numerical simulations show that the outcome is the development of *spiral arms* induced by the self-gravity within the disk. Gravitational forces set up by the spiral arms act much like magnetic tension in an MRI-unstable disk, and work to transport angular momentum outward and mass inward.

Gammie (2001) has shown that a disk is bound to fragment if the cooling time $\tau_{\mathrm{cool}} \stackrel{<}{\sim} 3\,\Omega^{-1}$. Figure 9.15 shows numerical simulations using SPH techniques for a disk of mass $M_{\mathrm{disk}} = 0.1\,M_\odot$– so a disk that is suffering from gravitational instabilities – in which the cooling time is long enough to prevent fragmentation and in which a spiral wave structure forms, and a disk that cools so rapid that fragmentation occurs. The simulation uses 250 000 SPH particles. For converting clumped regions in a sink particle the nearest 50 SPH particles

are checked to determine if they are gravitationally bound. For a $1\,M_\odot$ central star and a disk mass of $0.1\,M_\odot$, the individual fragments represent at least $6.7\,M_\oplus$ (earth masses) in gas and dust. These fragments tidally interact with the disc, heating it and causing $Q$ to increase to a value greater than unity. Once $Q > 1$ fragmentation ceases, and the disk, at that radius, becomes gravitationally stable. There is likely to be a significant amount of angular momentum transport driven by tidal interactions between fragments and the gaseous disc.

So, what is the cooling time for a circum-stellar disk? Let us assume that the disk is smooth (i.e. it has not fragmented). In that case viscous processes heat the disk at a local rate (see Eq. 9.35)

$$Q_{\text{heating}} = \frac{9}{8}\,\nu\,\Sigma\,\Omega^2. \tag{9.65}$$

This heating must be balanced by cooling. If the disk is optically thick, the local cooling rate from the disk surface is

$$Q_{\text{cooling}} = \sigma T_{\text{disk}}^4. \tag{9.66}$$

Using Eqs. (9.57) and (9.26) we may write the viscosity in the form

$$\nu = \alpha\,\frac{a^2}{\Omega} = \frac{\alpha}{\Omega}\,\frac{k\,T}{\mu\,m_{\text{p}}}, \tag{9.67}$$

realizing that $\alpha$ represents a 'gravitational viscosity', as it is induced by 'gravitational turbulence'. We may then define a typical cooling time

$$\tau_{\text{cool}} \;\; = \;\; \frac{U}{2Q_{\text{cooling}}} = \frac{U}{2Q_{\text{heating}}} = \frac{4}{9\gamma(\gamma-1)}\,\frac{1}{\alpha\,\Omega}, \tag{9.68}$$

where $U = a^2\Sigma/\left[\gamma\,(\gamma-1)\right]$ is the two-dimensional internal energy of the disk per unit area and $\gamma$ the two-dimensional adiabatic index (see Gammie 2001). The factor 2 reflects that the disk has two sides. We have set $Q_{\text{heating}} = Q_{\text{cooling}}$. The important thing is that the cooling time thus becomes very short if the orbital period $2\pi/\Omega$ is short and if the Shakura-Sunyaev $\alpha$ parameter is large. The former explains why instabilities first develop in the inner disk. For a di-atomic gas $\gamma = 7/5$ (see Eq. 3.5) one can only avoid fragmentation if $\alpha \stackrel{<}{\sim} 0.1$, as otherwise $\tau_{\text{cool}} \stackrel{<}{\sim} 3\Omega^{-1}$.

For quiescent T Tauri discs, $\alpha$ is estimated to be of the order of $10^{-2}$ or smaller. This would suggest that such disks are comfortably stable against fragmentation. The masses of these disks are of the order of $\sim 10^{-2}\,M_\odot$. However, some disks are considerably more massive. Massive disks may generate strong gravitational turbulence, hence have large $\alpha$, characterised by large density contrasts. The strong density contrasts result in larger torques due to self-gravity. If the cooling is too rapid, the disk cannot be stabilised against collapse, and fragmentation ensues. This may thus be a route toward giant planet formation, albeit likely not the main route.
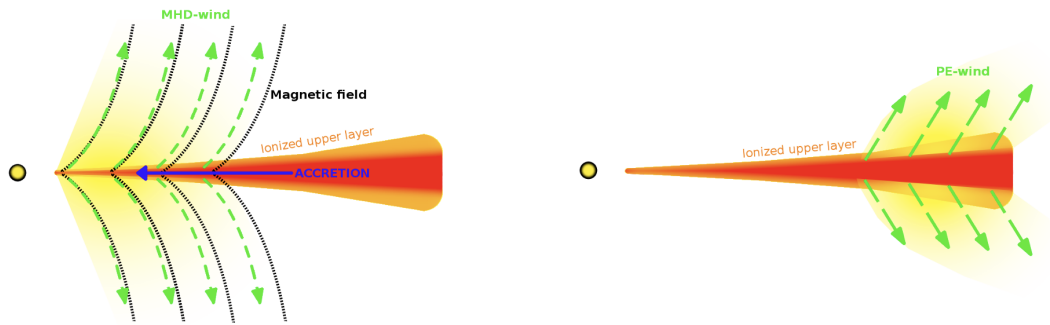
Figure 9.16: *Both magneto hydrodynamic (MHD) disk winds and photo-evaporative (PE) disk winds rely on EUV and X-ray photons, either from the central star or from the ambient medium, to create a hot (∼$10^4$ K) ionised disk surface layer. For MHD-winds, the differentially rotating Keplerian disk winds-up the magnetic field, creates a gradient in the disk magnetic field, allowing for the ions to reach the escape velocity following the magnetic field lines. In contrast to the PE-wind, this mechanism also extracts angular momentum form the disk (and thus drives accretion). MHD-driven winds are most effective in the inner parts of the disk. In PE-driven disk winds, the ions can escape in a thermal wind at distances where the local sound speed exceeds the local escape velocity $v_{esc} \propto r^{-1/2}$. Figure from: Per-Gunnar Valegård.*

## 9.6 Disk dispersal

Depletion of the gaseous proto-planetary disk due to accretion onto the central star is predicted to be a gradual process. If we consider the self-similar solution Eq. (9.54), for example, we find that for the adopted time-independent viscosity scaling with radius, $\nu \propto r$, the late-time behaviour of the surface density is

$$\Sigma \propto t^{-3/2}. \tag{9.69}$$

A disk that evolved as a result of accretion only would steadily become optically thin over an increasing range of radii, in the process transitioning rather slowly from a Class II to a Class III object. Observationally this is *not* what is observed. Although a number of disks intermediate in properties to Class II and III has been found – typically YSOs that lack a near-IR excess despite the presence of a robust mid-IR excess – the scarcity of such systems suggests that the dispersal phase of proto-planetary disks lasts of the order of $10^5$ yr (Simon & Prato 1995, Wolk & Walter 1996). The relative brevity of the dispersal timescale suggests that additional physical processes beyond viscous evolution contribute to the loss of gas from the disk.

Disk evolution can also be influenced by other dynamical processes such as disk winds and jets driven by stellar magnetic fields, dynamical encounters with nearby stars and photo-evaporation winds by either the central star or an external source (see Fig. 9.16 for the concepts of magnetically driven winds and photo-evaporation driven winds) . Hollenbach et al. (2000) discuss all these processes in detail and conclude that for most systems, viscous evolution and photo-evaporation are the dominant processes that shape the overall disk evolution. The other

processes mentioned can dominate over short timescales or in small regions of the disk (such as magnetically-launched jets that may operate very close to the star), or can be important for a small fraction of stars (such as tidal stripping by nearby stars, which affects $\sim 10\%$ of the stars in the Orion nebula cluster), but do not dominate the evolution of most T Tauri disks.

The following sections present a summary of photo-evaporation by the central star or an external source and briefly describe how disk evolution proceeds in the presence of photo-evaporation.

## Photo-evaporation

In the very surface layers of the disk the gas temperature can become very high, greatly exceeding the dust temperature. The warm surface gas can flow off the disk and escape the gravity of the star. Since the heating process responsible for these high temperatures is the radiation from the central star or a nearby O-star, this process is called *photo-evaporation*. The viscous evolution (i.e. accretion and spreading) of the disk, discussed in chapter 9.3, can be strongly affected by this photo-evaporation process and may significantly shorten the lifetime of a disk compared to pure viscous evolution. Photo-evaporation can also create inner holes or truncate the outer disk. This has relevance to observations of such disks, such as the percentage of young stars with infrared excess versus their age, or the inferred large inner holes of some disks. It has also far-reaching consequences for the formation of planets. We will discuss here first photo-evaporation by the central star and then turn to external sources.

The details of photo-evaporation depend upon the source of the irradiating flux – which can be the central star or external stars in a cluster – and the energy of the photons. Extreme ultraviolet (EUV) radiation, with energies $E > 13.6$ eV or $\lambda < 912$ Å, ionizes hydrogen atoms, producing a layer of hot gas whose temperature – around $10^4$ K – is almost independent of the density of the disk at the radii under consideration. Far ultraviolet (FUV) radiation, with energies $6$ eV $< E < 13.6$ eV dissociates $H_2$ molecules, creating a neutral atomic layer whose temperature depends upon the precise balance of heating and cooling processes. Typical values are 100–5000 K. The FUV photo-evaporation layer penetrates deeper into the disk than does the EUV layer. Here we mainly focus on disk photo-evaporation due to EUV irradiation by the central star, as it is this regime that is probably most important for the loss of gas from the planet-forming region of the disk.

## EUV photo-evaporation

Whether the EUV heating is sufficient to drive an evaporative flow depends on how the resulting thermal speed (or sound speed) compares to the local escape speed from the gravitationally bound system. A characteristic radius for thermal evaporation is the *gravitational radius* $r_g$ (see Fig. 9.17), where the sound speed (Eq. 3.3) equals the local escape speed
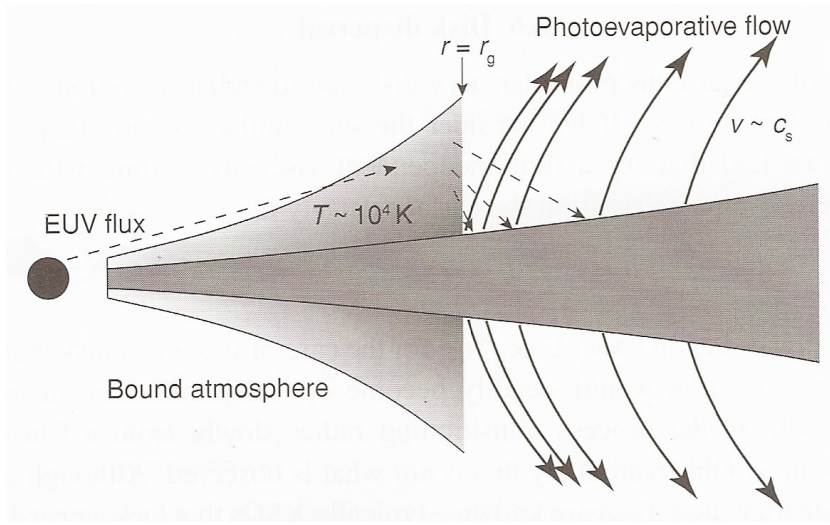
Figure 9.17: *The geometry assumed in simple models of disk photo-evaporation by a central stellar EUV source (Hollenbach et al. 1994). The EUV flux incident on the disk ionizes the surface layers, heating them to a temperature of around $10^4$ K. Inside a critical radius $r_\mathrm{g}$ the ionized layer forms a bound disk atmosphere that absorbs the stellar flux and re-radiates recombination radiation. At larger radii the gas escapes as a thermally driven disk wind. From: Armitage 2010, Astrophysics of planet formation.*

$v_{\mathrm{esc}} = (2GM/r)^{1/2}.$

$$r_\mathrm{g} = \frac{2GM_\star}{a^2} = \frac{2GM_\star\,\mu\,m_\mathrm{p}}{kT} \sim 15 \text{ AU} \left(\frac{T_{\mathrm{gas}}}{10^4 \text{ K}}\right)^{-1} \left(\frac{M_\star}{M_\odot}\right), \qquad (9.70)$$

for $\mu = 0.7$. Early analytic models made the simple assumption that photo-evaporation occurred for $r > r_\mathrm{g}$, and that the warm surface was gravitationally bound for $r < r_\mathrm{g}$. However, a closer look at the gas dynamics shows that this division happens not at $r_\mathrm{g}$ but at about $0.1 - 0.2\,r_\mathrm{g}$ (Liffman 2003, Adams et al. 2004, Font et al. 2004), and that this division is not entirely sharp. In other words, photo-evaporation happens mostly outside of the critical radius $r_{\mathrm{cr}} \sim 0.15\,r_\mathrm{g}$, though a weak evaporation occurs inside of $r_{\mathrm{cr}}$.

With $T_{\mathrm{gas}} \sim 10^4$ K, the critical radius for EUV-induced photo-evaporation is $r_{\mathrm{cr}} \sim 1 - -2\,(M_\star/M_\odot)$ AU. There is no fixed $r_{\mathrm{cr}}$ for FUV-evaporation because the FUV-heated gas has temperatures that depend on FUV flux and gas density, i.e., on $r$ and $z$. Therefore, $r_{\mathrm{cr}}(\mathrm{FUV})$ depends on $r$ and $z$, and may range from $3 - 150$ AU for solar mass stars.

Hollenbach et al. (1994) first outlined the essential physics of EUV-induced flows by the central star and presented an approximate analytic solution to the total mass-loss rate for a disk larger than $r_\mathrm{g}$. It is estimated to be

$$\dot{M}_{\mathrm{EUV}} \sim 1.6 \times 10^{-10} \left(\frac{\Phi_{\mathrm{EUV}}}{10^{41} \text{ s}^{-1}}\right)^{1/2} \left(\frac{M_*}{M_\odot}\right)^{1/2} \qquad (9.71)$$
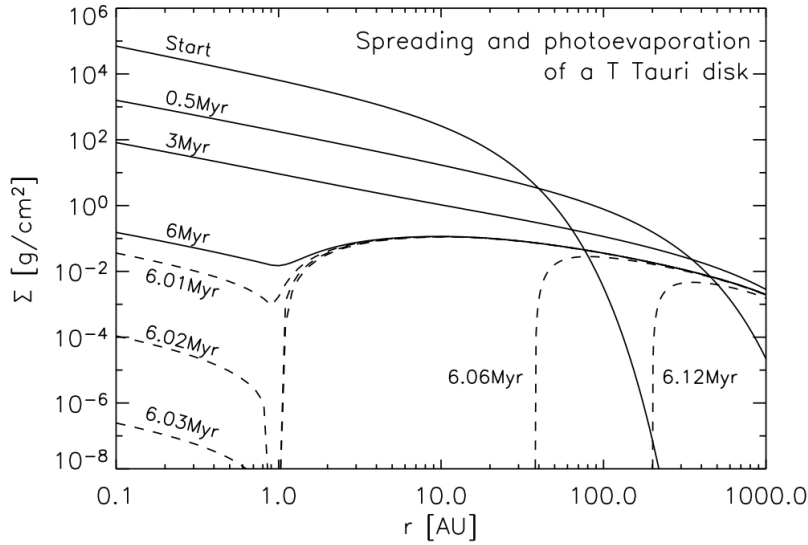
Figure 9.18: *The evolution of the surface density within a proto-planetary disk model that includes photo-evaporative mass loss. The mass of the star is $1\,M_\odot$ and the initial disk mass $0.05\,M_\odot$. The model assumes a viscosity $\nu \propto r$. No photo-evaporative mass loss is assumed to occur within $r_{\rm cr}$. The disk evolves slowly on Myr timescales until the stellar accretion rate drops to $\dot{M} \sim \dot{M}_{\rm wind}$, at which point the inner disk becomes cut off from the outer disk and drains rapidly (on a $10^5$ yr timescale) on to the star. From: Adapted from Alexander et al. 2006 (MNRAS 369, 229).*

in units of $M_\odot$/yr. $\Phi_{\rm EUV}$ is the total number of photons at wavelengths shortward of $912\,\text{Å}$ that the star supplies each second.

For an ionizing photon flux of $\Phi \sim 10^{41}\,{\rm sec}^{-1}$ the predicted mass-loss rate from the disk is essentially negligible at early times (when the stellar accretion rate is three orders of magnitude greater), but starts to dominate after a few viscous times of disk evolution. This behavior is qualitatively consistent with that required to reproduce the observed evolution of T Tauri disks, provided that the low-mass stellar EUV flux – which must be two to three orders of magnitude in excess of the present-day output of the Sun for the model to work – is high enough. Although there is a dearth of direct measurements, such an enhanced EUV output for young stars seems likely given the observed strength of other activity indicators (for example the X-ray flux).

## Photo-evaporation and viscous evolution

Incorporating photo-evaporation into time-dependent models of proto-planetary disk evolution is straightforward. The flow away from the disk carries the same specific angular momentum as the disk at the launch point, and hence the effect of photo-evaporation can be

captured as a simple mass sink in the disk evolution equation (9.15), i.e.

$$\frac{\partial \Sigma}{\partial t} = \frac{3}{r}\frac{\partial}{\partial r}\left[\sqrt{r}\,\frac{\partial}{\partial r}\left(\nu\,\Sigma\,\sqrt{r}\right)\right] + \dot{\Sigma}_{\mathrm{wind}}(r,t). \tag{9.72}$$

The mass loss term must be specified via a model of the photo-evaporation process.

Let us consider a disk that is actively accreting onto the star. In general, if the photo-evaporation drills a hole somewhere in the disk or Ôeats Õ its way from outside in, the forces of viscous spreading tend to move matter toward these photo-evaporation regions, which can accelerate the dissipation of the disk. If the disk has a steady accretion rate $\dot{M}$, then a gap forms once $\dot{M}_{\mathrm{wind}} \sim r^2\,\dot{\Sigma}_{\mathrm{wind}}$ exceeds $\dot{M}$. Since $r^2\,\dot{\Sigma}_{\mathrm{wind}} \sim r^{-0.5}$ for EUV photo-evaporation beyond $r_{\mathrm{cr}}$, the gap first forms at the minimum radius ($\sim r_{\mathrm{cr}}$) and then works its way outward.

The simulation shown in Fig. 9.18 starts from a given disk structure of about 0.05 $M_\odot$ (marked with ÔStartÕ in the figure). Initially the disk accretes and viscously spreads (solid lines). At $t = 6 \times 10^6$ yr the photo-evaporation starts affecting the disk. Once the EUV photo-evaporation has drilled a gap in the disk at $\sim 1$ AU, the destruction of the disk goes very rapidly (dashed lines). The inner disk accretes onto the star on an inner disk viscous timescale of $\sim 10^5$ yr. Once the inner hole is produced, the diffuse flux from the atmosphere of the inner disk is removed and the attenuation of the direct flux by this same atmosphere is also removed. This enhances the EUV photo-evaporation rate by the direct EUV flux from the star, and the effect magnifies as the inner hole grows as $\dot{M}_{\mathrm{wind}} \sim r_{\mathrm{inner}}^{1/2}$ (which can be derived from applying a simple Strömgren criterion). This leads to a rapid erosion of the outer disk from inside out. In this model the disk viscosity spreads to $> 1000$ AU; however, FUV photo-evaporation (not included) will likely truncate the outer disk.

The disk evolution can be split into three phases: (1) the viscous phase which lasts a few Myr, (2) the gap-opening phase which is shorter than $10^5$ yr and (3) the disk clearing phase which lasts a few $10^5$ yr. During the viscous evolution phase, most of the disk mass is accreted onto the star and the disk spreads to large radii. Photo-evaporation is negligible. During the second phase, the photo-evaporative wind drills a gap into the disk. This stops the mass supply from the outer disk, and the inner disk accretes onto the star on its viscous timescale. During the clearing phase (third phase), the inner disk is optically thin to irradiating and ionizing photons and evaporates from inside out.

## External evaporation

In crowded regions of star formation, especially in the presence of high mass stars, external photo-evaporation can also play a significant role. This effect has been observed, for instance, for the proto-planetary disks in the Orion nebula (Fig. 9.19, O'Dell et al. 1993). These disks around low mass stars display cometary-shaped nebulae that point away from the illuminating massive star, hence are being photo-evaporated by the nearby bright O-type stars. The evaporating gas is bright in typical nebulae emission lines such as H $\alpha$ and [N II]. The photo-evaporative flow is mainly driven by the non-ionizing (hydrogen) FUV photons.
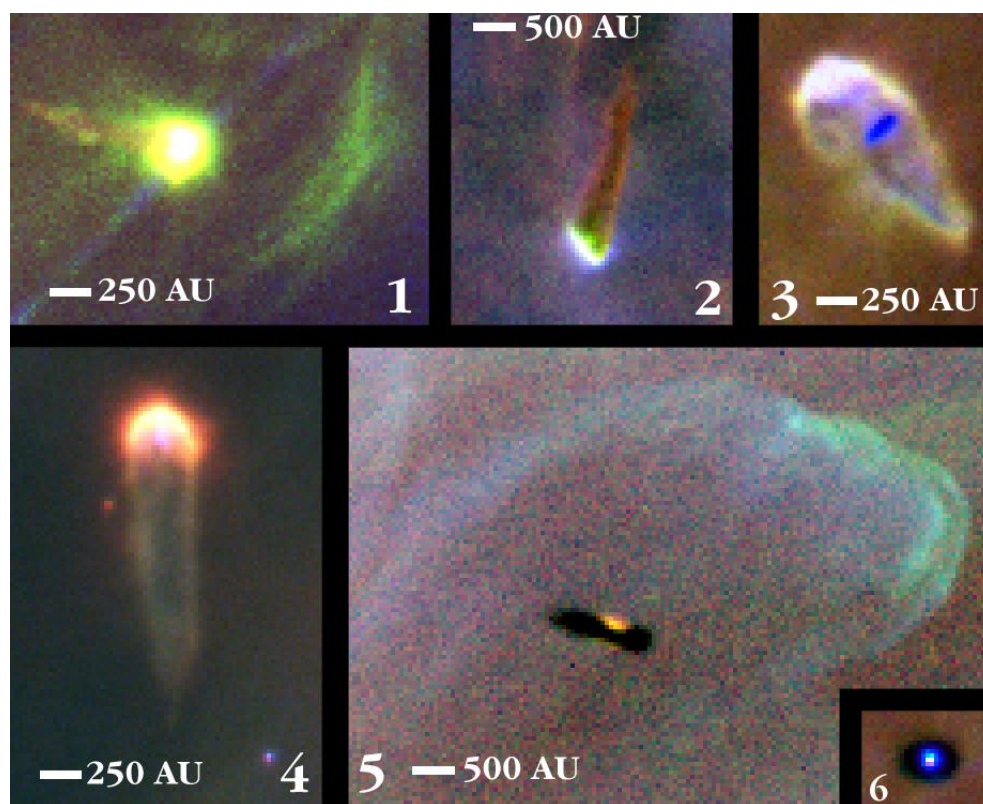
Figure 9.19: *Hubble Space Telescope images of proto-planetary disks in Orion, in the process of being photo-evaporated by nearby bright O-type stars. From: O'Dell et al. (1993). The bright part of the disk is due to gas that is ionized by the high-energy photons of nearby massive stars. Ionized proto-planetary disks are often referred to as* proplyds. *The black part is tracing the bulk of the disc, where the dust makes it optically thick and thus visible in absorption against the bright background.*

The cometary-shaped nebulae are explained by a combination of the photo-evaporative flow from the disk surface and the dynamic interaction with the O star winds. Typical evaporation mass-loss rates are of the order of $10^{-7}$ M$_\odot$/yr, but can differ substantially depending on the distance from the illuminating source.

Even though one might be led to the conclusion that this strong photo-evaporation leads to extremely short lifetimes for disks, this is not necessarily true. The gravitational radius for FUV photo-evaporation is much larger than that for EUV photo-evaporation, because the flow is driven at substantially lower temperatures, typically below $1000$ K (see Eq. 9.70). Since material inside $\sim 0.15\, r_\mathrm{g}$ is mostly bound, disks of that size can survive over much longer timescales.

## Exercise 9.1

We consider a particle of mass $m$ that is in orbit around a star of mass $M$, moving at the Keplerian orbital velocity $v_\phi$ (Eq. 9.1).

a)  Show that the particle is in virial equilibrium, that is, that it fulfills the relation $2E_\text{kin} + E_\text{pot} = 0$.

b)  Evil space aliens suddenly steal part of the mass of the star. How much of the mass do they need to steal for the particle $m$ to become unbound to the star?

## Exercise 9.2

In this exercise we investigate the circumstances in which the gravitational force of the disk itself is more important for the disk structure than the gravitational force of the central star. We first derive the gravitational acceleration at a point $P$ some distance $z$ above a razor thin disk. Point $P$ is located on the axis of symmetry (see figure). The disk has a constant surface density $\Sigma$. From point $P$ the disk can be seen up to an angle $\alpha$ measured from the axis of symmetry.
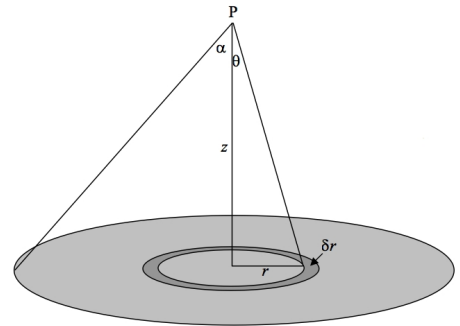


a)  Give the contribution to the gravity $dg_\text{disk}$ from an elemental annulus at radius $r$ of thickness $dr$ as a function of $\Sigma$, $z$ and $r$.

b)  Integrate over all annuli from $\theta$ is 0 to $\alpha$ and show that the gravity is given by

$$g_\text{disk} = 2\pi G \Sigma \left(1 - \cos\alpha\right). \tag{9.73}$$

Hint: do yourself a huge favor and integrate with respect to $\theta$, where $r = z \tan\theta$, such that $dr = (z/\cos^2) \cdot d\theta$. Verify that for an infinitely large disk $g$ is given by Eq. (9.28).

c)  The disk that we consider is large (such that for practical purposes $\alpha \to \pi/2$). We assume that self-gravity by the disk dominates over stellar gravity when $g_\text{disk} > g_{z,\star}(z = h)$ at least at the outer disk radius. $g_{z,\star}(z = h)$ is the vertical component of the stellar gravity at a height $z = h$ above the disk. Show that this implies that the disk becomes unstable for its own gravity if

$$\frac{M_\text{disk}}{M_\star} > \eta, \tag{9.74}$$

where $M_\text{disk}$ is the total disk mass and $M_\star$ is the stellar mass. Derive an expression for $\eta$ as a function of $h$ and $r$.

d)  Discuss what may happen if the disk becomes unstable due to self-gravity.

**Exercise 9.3**

Evil space aliens compress the Earth to an infinitely thin disk of uniform surface density $\Sigma_\oplus$ without changing its radius. What is the gravitational acceleration experienced by a survivor standing on the surface of this disk?

**Exercise 9.4**

Derive diffusion equation (9.56) from the surface density evolution equation (9.15).

**Exercise 9.5**

a) Show that the local cooling time of a viscously heated two-dimensional Shakura-Sunyaev disk with two-dimensional adiabatic index $\gamma = 7/5$ is given by

$$\tau_{\text{cool}} \simeq \frac{0.1}{\alpha} P, \tag{9.75}$$

where $P$ is the local Keplerian period.

b) In section 9.5 we have seen that a viscous disk prone to gravitational collapse (so, fulfilling Eq. 9.64) can only avoid fragmentation if $\alpha \lesssim 0.1$. So, equivalently: a massive gravitational unstable viscous disk can *not* avoid fragmentation if the local cooling time is longer / shorter than the local orbital period. Longer or shorter?

# 10

# Building planets (part I) - growth by aggregation

Gravitational instabilities in massive circum-stellar disks may be a route toward planet formation (see Sect. 9.5). An alternative mechanism is that of the collision of initially sub-micron sized solid state particles – present already in the molecular cloud – and subsequent growth via aggregation. In the remaining three chapters of these lecture notes we give a conceptual overview of the different steps that must lead from an ensemble of such sub-micron sized particles to a modest number of rocky, terrestrial sized planets and/or gas-giants orbiting the central star.

The entire process of planet formation spans around 40 orders of magnitude in mass and 14 orders of magnitude in radius (see Fig. 10.1). This growth range is usually subdivided into a number of stages, in which different physics dominate the growth of the particles (e.g. Chiang & Youdin 2010; Morbidelli et al. 2012). In presenting these stages we closely follow Mulders (2013); see also figure 10.2:

(1) Growth by aggregation. In the initial step of planet formation microscopic dust grains grow through aggregation following mutual collisions. Due to the high densities in proto-planetary disks, grains collide frequently and can grow much larger than in the interstellar medium or molecular clouds. The efficiency with which grains adhere upon collision – loosely speaking how 'sticky' they are – plays an important role in the process. They grow quickly from micron to millimeter size (Weidenschilling 1980) – in $10^3 - 10^4$ yr within a few AU of the star –, however, as grains grow larger, the relative velocities of the colliding components keeps increasing, until the impact velocities are so high that collisions become destructive. This is called the *fragmentation barrier* and provides an upper limit to the size grains can grow by aggregation, ranging from roughly a meter in size in the inner disk to a millimeter in size in the outer disk[1].

---

[1]There are other barriers along the way as well, such as the bouncing barrier (Zsom et al. 2010) – related to a decreased sticking efficiency at larger sizes – and the electrostatic barrier (Okuzumi 2009), which prevents collisions between charged dust grains. These barriers may temporarily stall the planet formation process, but are
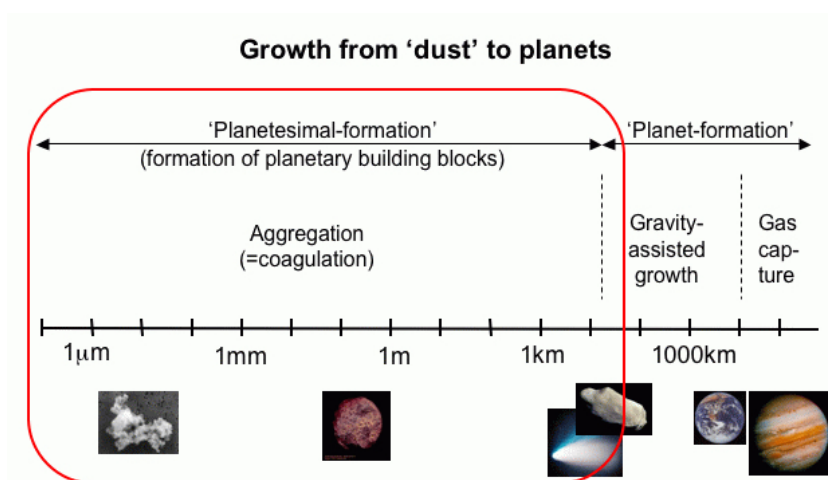
Figure 10.1: *Approximate size range spanned in the planet formation process. The images shown below the scale bar are objects found in the solar system, from left to right: an interplanetary dust particle, a meteorite, a comet and an asteroid, a terrestrial planet (Earth) and a gas giant planet (Jupiter). From: Kees Dullemond.*

(2) The intermediate regime. Growth of dust grains is most difficult to understand in the regime where grains are too large to grow by aggregation (Blum & Wurm 2008), but too small to grow by gravitational interaction. Apart from reaching the fragmentation barrier, particles growing to a size of about a meter experience a strong headwind from the gas causing them to drift inwards, quickly (in less than $10^5$ yr) depleting the disk of solids to form planets (Weidenschilling 1977). This barrier is called the meter-sized barrier or *radial-drift barrier*. A promising mechanism to cross this regime and overcome these barriers is *particle concentration*. Two centimeter-sized particles do not gravitationally attract each other, but a quadrillion might. Long-lived structures in the gas such as pressure bumps, turbulent eddies (vortices) or dead zones can trap large numbers of particles, making them susceptible to gravitational collapse directly into larger bodies (e.g. Johansen et al. 2006) before they can drift or fragment. The size distribution of asteroids in the solar system indicates that they indeed formed in this way, bridging directly the regime from centimeter-sized grains to kilometer sized asteroids (Morbidelli et al. 2009). Still, the open question in general for this class of models remains whether the local particle density can reach the large values needed to trigger collapse.

(3) Gravitational growth. Once kilometer-sized bodies, referred to as *planetesimals*[2], have been formed, gravity takes over as the main driver of growth. At these sizes, collisions do not lead to fragmentation, as their gravitational pull is strong enough to retain most

---

surmountable. They are important because they keep a small fraction of micron-sized dust in these disks that is observed in proto-planetary disks at all ages.

[2]Be ware not to take the size definition of planetesimals too strict: they can refer to anything from 1–100 km scale bodies.
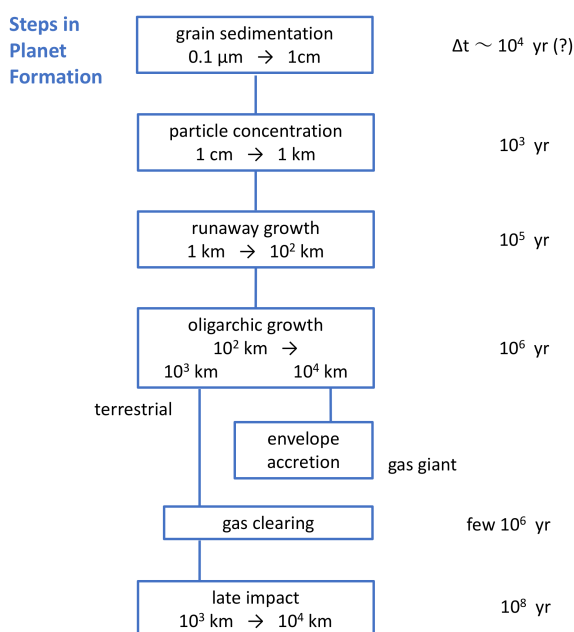
Figure 10.2: *Different stages in planet building. The required time intervals als indicated at the right. For grain sedimentation the quoted time is a lower limit, based on the simplest, quiescent disk model.*

of the material of a collision in a single body, i.e. a collision may shatter the collision partners but gravitational forces will bring together the bulk of the debris in a single (or binary) object. The gravitational growth phase itself can be further divided in two phases. In the first phase, more massive planetesimals have a higher chance of colliding. They grow faster than the smaller one, and this phase is therefore called *runaway growth* (Greenberg et al. 1978). In the second phase, the gravity of these largest bodies prevents the growth of smaller planetesimals altogether. They continue to grow themselves by slowly capturing the less massive planetesimals. This phase is called *oligarchic growth* (Kokubo & Ida 1998), and creates planetary embryos of about a thousand kilometers, the size of dwarf planets such as Ceres and Pluto.

(4) Final stages of assembly. In the final stages of planet formation, the proto-planetary disk is seeded with planetary embryos or proto-planets that have collected most of the solid materials (i.e. the dust) in their vicinity. They are too tightly packed to form a stable planetary system, and will continu to collide and grow. Whether these embryos will form *terrestrial planets* or *gas giants* depends on a number of factors, most importantly their mass and the available amount of gas. If a proto-planet reaches a mass of about 10 Earth masses while the disk still has plenty of gas, it can accrete significant amounts of this gas to become a gas giant like Jupiter in our solar system (Pollack et al. 1996). If a 10 $M_\oplus$ object is reached after most of the gas has been dispersed, an object with the mass of Neptune or a super-Earth will remain. Cores that never reach 10 $M_\oplus$ will form terrestrial planets. They will slowly assemble in the years after all the gas has been

dispersed, until a stable planetary system remains (or not). Overall, the core accretion models provide a good match to the observed population of exoplanets (Mordasini et al. 2012).

## 10.1 Grain growth and dust settling

Grain particles in a proto-planetary disk moving with a velocity $v$ *relative to the gas* will suffer *aerodynamic drag*. Aerodynamic drag on particles is important for understanding both the vertical distribution and radial motion of dust and larger bodies within disks. We discuss here some fundamental aspects of drag. Figure 10.3 provides a simplistic overview of the dynamics of dust particles in proto-planetary disks.

Consider a grain of radius $s$ and material density $\rho_{\mathrm{m}}$. We calculate the aerodynamic force experienced by the particle when it moves at a velocity $v$ *relative to the local velocity of the gas disk*. In calculating the force one may identify two physical regimes. If the size $s$ of the particle is small relative to the mean-free path $\ell$ of gas molecules within the disk, then the gas on the scale of the particle is effectively a collision-less ensemble of molecules with a Maxwellian velocity distribution. The drag force in this regime – called *Epstein drag* – is normally the most relevant for small particles within proto-planetary disks and we focus on this regime in our discussion. In the alternative regime – i.e. $s \gtrsim \ell$ – the disk gas behaves as a fluid, flowing around the obstruction presented by the particle. The drag associated with this situation is referred to as *Stokes drag*.

In either regime the drag force scales with the frontal area $\pi s^2$ that the particle presents to the gas. This means that the acceleration $g_{\mathrm{drag}}$ caused by gas drag – which is proportional to the drag force divided by the particle mass – *decreases* with particle size as $g_{\mathrm{drag}} \propto s^{-1}$ and eventually becomes negligible once bodies of planetesimal size have formed.

### Drag force

We derive the Epstein drag by considering the frequency $f$ of collisions between the particle and gas molecules, which is the product of the collision cross-section, relative velocity $v$, and molecular number density $n = \rho/\mu\, m_{\mathrm{H}}$, where $\rho$ is the gas density and $\mu$ the mean molecular weight. Within the gas the mean thermal speed of the molecules is $v_{\mathrm{th}}$. For a Maxwellian distribution of temperature $T$ this velocity is

$$v_{\mathrm{th}} = \sqrt{\frac{8kT}{\pi\mu m_{\mathrm{H}}}} = \sqrt{\frac{8}{\pi}}\, a, \tag{10.1}$$

where $a$ is the sound speed.

1. Turbulence distributes particles vertically and radially
2. Growing particles sediment to the mid-plane of the disk
3. In the outer disk, particles of mm sizes drift quickly inwards
4. Particle collisions can have many different outcomes
   a) Sticking (= growth)
   b) Bouncing (= growth neutral)
   c) Fragmentation with net growth of the larger body
   d) Fragmentation

5. Volatile species condense or form on particle surfaces
   Particles sediment to the disk mid-plane and drift inward
6. In the hotter regions, volatiles are released back into the gas phase
7. Small fragments can be turbulently mixed up to the disk surface
   and can be carried away by disk winds
8. Planets or other effects can accumulate larger particles

Figure 10.3: *Illustration of the structure and grain dynamics processes in proto-planetary disks. On the left side the main grain transport and growth/destruction mechanisms of refractory species (in orange) are depicted. The different lengths of the arrows denote the different velocities of small and large grains. On the right side the formation and destruction of volatile species is shown as well as some possible ultimate fates of refractory grains. From: Til Birnstiels homepage.*

Up to factors of the order of unity, the frequency with which gas molecules collide with the 'front' side of the particle is

$$f_+ \approx \pi s^2 \left( v_{\text{th}} + v \right) \frac{\rho}{\mu\, m_{\text{H}}}, \tag{10.2}$$

while the collision frequency on the back side is

$$f_- \approx \pi s^2 \left( v_{\text{th}} - v \right) \frac{\rho}{\mu\, m_{\text{H}}}. \tag{10.3}$$

The momentum transfer $\Delta p$ per (head-on or rear-on) collision is the sum of the impact kick and recoil kick momentum transfer, i.e.

$$\Delta p = \mu\, m_{\text{H}} \left( v + v_{\text{th}} \right) - \mu\, m_{\text{H}} \left( v - v_{\text{th}} \right) = 2\mu\, m_{\text{H}}\, v_{\text{th}}, \tag{10.4}$$

which implies that the net force in the Epstein regime scales as

$$F_{\text{drag}} = -\Delta p \left( f_+ - f_- \right) = -4\pi \rho\, s^2\, v_{\text{th}}\, v, \tag{10.5}$$

where the minus sign signifies that the particle is decelerated by the gas. The force is linear in the relative velocity and proportional to the surface of the particle and to the thermal speed of

molecules in the gas. A more accurate derivation, valid for $s < \ell$, $v \ll v_{\text{th}}$ and a Maxwellian distribution of molecular speeds, yields

$$\mathbf{F}_{\text{drag}} = -\frac{4\pi}{3} \, \rho \, s^2 \, v_{\text{th}} \, \mathbf{v}. \tag{10.6}$$

The drag force, of course, acts in the opposite direction to the vector $\mathbf{v}$ describing the relative velocity between the particle and the gas.

### 10.1.1  Dust settling

We consider the vertical settling of dust particles to the disk mid-plane in the absence of turbulence. To begin with, we quantify the coupling between the solid and gas components of the disk by defining the *friction timescale* for a particle of mass $m$ as

$$\tau_{\text{fric}} = \frac{m \, v}{|\mathbf{F}_{\text{drag}}|}. \tag{10.7}$$

The friction timescale measures the time in which drag modifies the relative velocity $v$ significantly. Writing $m = (4/3) \, \pi \, s^3 \, \rho_{\text{m}}$, $\tau_{\text{fric}}$ takes the simple form in the Epstein drag regime

$$\tau_{\text{fric}} = \frac{\rho_{\text{m}}}{\rho} \, \frac{s}{v_{\text{th}}}. \tag{10.8}$$

Adopting $\rho = 10^{-9} \, \text{g cm}^{-3}$ and $v_{\text{th}} = 10^5 \, \text{cm s}^{-1}$, conditions appropriate for a particle at the mid-plane of a typical proto-planetary disk at $r = 1 \, \text{AU}$, we obtain for a particle of size $s = 1 \, \mu\text{m}$ and material density $\rho_{\text{m}} = 3 \, \text{gr cm}^{-3}$ a friction time $\tau_{\text{fric}} \approx 3 \, \text{s}$. Small dust particles are thus very tightly coupled to the gas.

We now consider the vertical forces acting on a small dust particle that is initially at rest at a height $z$ above the mid-plane. The $z$-component of the stellar gravity yields a downward force (see Eqs. 9.24 and 9.2)

$$F_{\text{grav,z}} = m \, \Omega^2 \, z, \tag{10.9}$$

where we have assumed that $z \ll r$. The *gas* in the disk is supported against this force by an upward directed pressure gradient, but no such force acts on a solid particle. The dust particle will accelerate, until, after about a friction timescale, a force balance is established between gravity and drag. From then on, the particle drifts toward the disk mid-plane with a (terminal) velocity given by equating Eqs. 10.6 and 10.9. This yields

$$v_{\text{settle}} = \frac{\rho_{\text{m}}}{\rho} \, \frac{s}{v_{\text{th}}} \, \Omega^2 \, z. \tag{10.10}$$

The proportionality to $z/\rho$ implies that in the absence of turbulence dust particles would be expected to settle out of the uppermost disk layers rather quickly. However, it will – strictly speaking – take an infinite amount of time to reach the mid-plane at $z = 0$ as $v_{\text{settle}} \propto z$. Relaxing on this, we approximate the settling time by

$$\tau_{\text{settle}} \simeq \int_0^h \frac{dz}{v_{\text{settle}}} \simeq \int_{h/10}^h \frac{dz}{v_{\text{settle}}} \simeq \frac{\rho}{\rho_{\text{m}}} \, \frac{v_{\text{th}}}{s} \, \frac{\ln(10)}{\Omega^2} = \frac{2 \ln(10)}{\pi} \, \frac{\Sigma}{\rho_{\text{m}} \, s \, \Omega} \tag{10.11}$$

In the third approximation we used that in the final scale height the density is almost constant (see the discussion below Eq. 9.26). Adopting the Gaussian density profile Eq. 9.25 (and 9.26 and 9.27) appropriate for a vertically isothermal disk, taking $\rho$ to be the mid-plane density, and noting that the mean speed differs from the sound speed that determines the vertical scale-height $h$ by only a numerical factor (see Eq. 10.1), we obtain the final expression for the settling time.

Inserting $\rho = 6 \times 10^{-10} \, \mathrm{g \, cm^{-3}}$, $z = 0.02 \, \mathrm{AU}$, and $v_{\mathrm{th}} = 10^5 \, \mathrm{cm \, s^{-1}}$, i.e. numbers roughly appropriate for a $1 \, \mu\mathrm{m}$ particle at $z \sim h$ at 1 AU from a solar mass star, one finds that $v_{\mathrm{settle}} \approx 0.06 \, \mathrm{cm \, s^{-1}}$. For the typical values given above $\tau_{\mathrm{settle}}$ is about $10^6$ yr.

### 10.1.2 Single particle settling with aggregation

The estimate of the dust settling time given above is incomplete because it ignores the likelihood that dust particles will collide with one another and adhere, hence grow during the settling process. The settling velocity increases with particle size, so any such aggregation hastens the collapse of the dust toward the disk mid-plane.

To estimate how fast particles could grow during sedimentation we apply a simple particle growth model devised by Safronov (1969). Imagine that a single 'large' particle of radius $s$ is settling toward the disk mid-plane at velocity $v_{\mathrm{settle}}$ through a medium of much smaller solid particles. By virtue of their small size the settling of small particles can be neglected. If every collision leads to aggregation the large particle grows in mass at a rate that reflects the amount of solid material in the volume swept out by its geometric cross-section

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \pi s^2 \, v_{\mathrm{settle}} \, f_{\mathrm{dg}} \, \rho(z),$$

(10.12)

where $f_{\mathrm{dg}}$ is the dust-to-gas ratio in the disk. Substituting for the settling velocity one finds

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \frac{3}{4} \frac{\Omega^2 \, f_{\mathrm{dg}}}{v_{\mathrm{th}}} \, z \, m.$$

(10.13)

Because $z = z(t)$ this equation cannot generally be integrated immediately, but rather must be solved in concert with the equation for the height of the particle above the mid-plane

$$-\frac{\mathrm{d}z}{\mathrm{d}t} = v_{\mathrm{settle}} = \frac{\rho_{\mathrm{m}}}{\rho} \frac{s}{v_{\mathrm{th}}} \Omega^2 \, z.$$

(10.14)

Solutions to these coupled equations provide a very simple model for particle growth and sedimentation in a non-turbulent disk.

Figure 10.4 shows numerical solutions to Eqs. (10.13) and (10.14) for initial particle sizes of $0.01$, $0.1$ and $1 \, \mu\mathrm{m}$. The particles settle from an initial height $5h$ through a disk whose parameters are chosen to be roughly appropriate to a (laminar, i.e. non-turbulent) Solar Nebula model at 1 AU from the sun. Both particle growth and vertical settling are found to be extremely
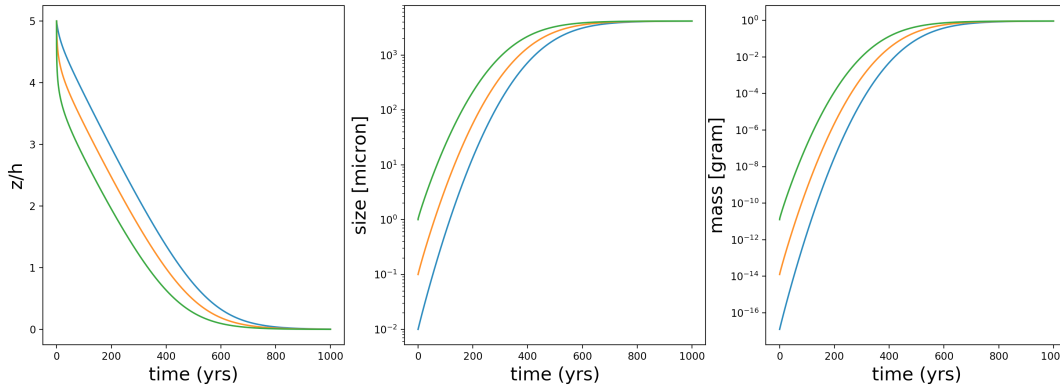
Figure 10.4: *The settling and growth of a single particle in a laminar (non-turbulent) proto-planetary disk. A single particle with initial size $s = 1\,\mu m$ (green line), $0.1\,\mu m$ (orange line), or $0.01\,\mu m$ (blue line), accretes all smaller particles it encounters while settling toward the disk mid-plane. The smaller particles are assumed to be at rest. The left panel shows the height above the mid-plane as a function of time, the middle and right panel the particle radius $s$ and mass $m$, respectively. For this example the disk parameters adopted are: an orbital radius $r = 1\,AU$ from a solar mass star, scale height $h = 0.02\,AU$, surface density $\Sigma = 10^3\,g\,cm^{-2}$, $f_{dg} = 0.01$, and a mean thermal speed $v_{th} = 10^5\,cm\,s^{-1}$. The dust particle is taken to have a material density $\rho_m = 3\,g\,cm^{-3}$ and to start settling from an initial height $z_o = 5h$. Based on: Armitage 2010. Astrophysics of planet formation.*

rapid. With the inclusion of aggregation, particles settle to the disk mid-plane on a timescale of the order of $10^3$ yr – more than two orders of magnitude faster than the equivalent timescale in the absence of particle growth. By the time that the particles reach the mid-plane they have grown to a final size of a few mm, irrespective of their initial radius.

The single particle model described above is very simple, both in its neglect of turbulence and because it assumes that the only reason that particle-particle collisions occur is because the particles have different vertical settling velocities. Other drivers of collisions include Brownian motion, turbulence, and differential radial velocities. The basic result, however, is confirmed by more sophisticated models (e.g. Dullemond & Dominik 2005), which show that, if collisions lead to particle adhesion, growth from sub-micron scales up to small macroscopic scales (of the order of a mm) occurs rapidly.

This means that there is no time scale problem associated with the very earliest phases of particle growth. Indeed, what is more problematic is to understand how the population of small grains – which are unquestionably present given the IR excesses characteristic of classical T Tauri stars (see Fig. 9.13) – persists to late times. The likely reason is *particle fragmentation* in sufficiently energetic collisions, which allows a broad distribution of particle sizes to survive out to late times. Fragmentation is not likely given collisions at relative velocities of the order of a few $cm\,s^{-1}$ – values typical of settling for micron-sized particles – but becomes more probable for collisions at velocities of $m\,s^{-1}$ or higher.

### 10.1.3  Settling in the presence of turbulence

After you leave your student room, suspended dust particles settle out of the air readily. When in your student room, the dust particles can remain suspended because of the vigorous air currents generated by you moving about industrially. The same physics applies within proto-planetary disks. Turbulence – probably but not necessarily the same as that responsible for angular momentum transport – acts to stir up small solid particles and this prevents them from settling into a thin layer at the disk mid-plane.

The conditions necessary for turbulence to stir up the dust enough to oppose vertical settling can be estimated by comparing the settling time Eq. (10.11) with the timescale on which diffusion will erase spatial gradients in the particle distribution. To diffuse vertically across a scale $z$ requires a timescale (see Eq. 9.55)

$$\tau_{\text{diffuse}} \approx \frac{z^2}{\nu},  \tag{10.15}$$

where we have assumed that the turbulence stirring up the particles is the same turbulence responsible for angular momentum transport within the disk. Using Eq. (9.57) for the viscosity $\nu$, we can compute the minimum value of $\alpha$ required for turbulence to oppose settling. We find at $z = h$

$$\alpha \gtrsim \frac{\pi}{2 \ln(10)} \frac{\rho_{\text{m}} s}{\Sigma},  \tag{10.16}$$

which is roughly the ratio between the column density through a single solid particle ($\rho_{\text{m}} s$) and that of the gas disk itself. For small particles this critical value of $\alpha$ is extremely small. If we take $\Sigma = 10^2 \, \text{g cm}^{-2}$, $\rho_{\text{m}} = 3 \, \text{g cm}^{-3}$, and $s = 1 \, \mu\text{m}$, for example, we obtain $\alpha \gtrsim 10^{-5}$. This implies that small particles of dust will remain suspended throughout much of the vertical extent of the disk in the presence of turbulence with any plausible strength. Only if we consider particles of radius 1 mm – a size that we argued might form very rapidly – we find that the critical value of $\alpha \sim 10^{-2}$. This value is compatible to most large scale estimates of $\alpha$ for proto-planetary disks. Particles of this size and above will therefore not have the same vertical distribution as the gas in the disks – they will settle in the mid-plane. This shows that substantial growth is required – by a factor $10^3$ in size or $10^9$ in volume – before settling takes place.

## 10.2  Radial drift of solid particles

Momentum equation (3.23) or (9.16) contains a force term due to the gradient in the gas pressure. The *gas* in the proto-planetary disk is therefore partially supported against gravity by an outward (radial) pressure gradient, and as a result orbits the star at a slightly sub-Keplerian velocity. Considering a stationary axisymmetric flow in which the gravitational potential is

dominated by that of the star, the radial component of the momentum equation implies that the orbital velocity of the gas $v_{\phi,\mathrm{gas}}$ is given by

$$\frac{v_{\phi,\mathrm{gas}}^2}{r} = \frac{G\,M}{r^2} + \frac{1}{\rho}\frac{\mathrm{d}p}{\mathrm{d}r}. \tag{10.17}$$

As the pressure near the disk mid-plane normally decreases outward, the second term on the right-hand-side is negative and the azimuthal velocity of the gas is slightly less than the Keplerian velocity Eq. (9.1) of a solid particle orbiting at the same radius (and that is not supported by the gas pressure gradient). In special cases, for instance near the inner edge of a dead zone (see Sect. 9.4), the gradient may be positive leading to local super-Keplerian motion of the gas. To quantify the difference we write the variation of the mid-plane pressure as a power-law

$$p = p_\circ \left(\frac{r}{r_\circ}\right)^{-n}, \tag{10.18}$$

where $p(r_\circ) = p_\circ$. Normally $n > 0$ (leading to sub-Keplerian gas motion). In special cases locally $n < 0$ (leading to local super-Keplerian gas motion). Using $p = \rho\,a^2$ (see Eq. 3.3), we find

$$v_{\phi,\mathrm{gas}} = v_{\mathrm{K}}\left[1 - n\left(\frac{a}{v_{\mathrm{K}}}\right)^2\right]^{1/2} = v_{\mathrm{K}}\left[1 - n\left(\frac{h}{r}\right)^2\right]^{1/2} \equiv v_{\mathrm{K}}\left[1 - \eta\right]^{1/2} \tag{10.19}$$

where we have used Eq. (9.26) and (9.2) to write down the second equality. The parameter $\eta = n\,(a/v_{\mathrm{K}})^2 = n\,(h/r)^2$. For a disk with a constant value of $h/r$, Eq (9.26) implies that $a \propto r^{-1/2}$. If we further assume a surface density profile $\Sigma \propto r^{-1}$, Eq (3.3) leads to $n = 3$. If we take $h/r = 0.05$, one obtains

$$v_{\phi,\mathrm{gas}} \simeq 0.996\,v_{\mathrm{K}}. \tag{10.20}$$

When considering the motions of the gas alone this difference is utterly negligible and we can safely assume the gas moves at the Keplerian velocity. The slightly lower gas velocity is however important for the solid particles in the gas, as it results in aerodynamic drag and resultant orbital decay.

For a small dust particle, aerodynamic coupling to the gas is very strong (because the friction time Eq. (10.7) is short relative to the orbital period $2\pi/\Omega$). To a good approximation the dust will be swept along with the gas, and its azimuthal velocity will equal that of the disk gas. Since this is sub-Keplerian, the centrifugal force will be insufficient to balance gravity (recall, it does not profit from the radial gas pressure gradient), and the particle will spiral inward with a certain radial velocity. Inward radial drift also occurs for rocks-sized particles that are poorly coupled to the gas (because $\tau_{\mathrm{fric}}$ is much longer than the orbital period). The rocks-sized objects orbit with an azimuthal velocity that is faster than the motion of the disk gas, and that is thus very close to the Keplerian speed. The faster motion implies that the rock experiences a 'headwind' that tends to remove angular momentum from the orbit. The loss

of angular momentum results in inward drift. So, though the physics is not identical, both the small and large particles spiral in towards the star.

We can now write down the equations of motion for the radial and azimuthal motions including the drag force, yielding expressions for $dv_r/dr$ and $dv_\phi/dr$, where $v_r$ and $v_\phi$ are the radial and azimuthal velocities of the particles

$$\frac{\mathrm{d}v_r}{\mathrm{d}t} = \frac{v_\phi^2}{r} - \Omega_K^2 \, r + g_{\mathrm{drag},r} = \frac{v_\phi^2}{r} - \Omega_K^2 \, r - \frac{1}{\tau_{\mathrm{fric}}} \left( v_r - v_{r,\mathrm{gas}} \right) \qquad (10.21)$$

$$\frac{\mathrm{d}(r\, v_\phi)}{\mathrm{d}t} = g_{\mathrm{drag},\phi} = -\frac{r}{\tau_{\mathrm{fric}}} \left( v_\phi - v_{\phi,\mathrm{gas}} \right), \qquad (10.22)$$

where $g_{\mathrm{drag}} = (v - v_{\mathrm{gas}})/\tau_{\mathrm{fric}}$ follows from Eq. (10.7). The latter equation describes the rate of change of specific angular momentum $j = r\, v_\phi$, being equal to the torque per unit mass $g_{\mathrm{drag},\phi} = G/m = r\, g_\phi$ (see Eq. 9.8). The solution to this set of equations can be obtained analytically when some justifiable simplifications are made and is, for instance, given in Armitage[3] (2010). One obtains an expression for the radial velocity of the particles

$$v_r = \frac{q_{\mathrm{fric}}^{-1} \, v_{r,\mathrm{gas}} - \eta \, v_K}{q_{\mathrm{fric}} + q_{\mathrm{fric}}^{-1}}, \qquad (10.23)$$

where $q_{\mathrm{fric}} = \tau_{\mathrm{fric}} \, \Omega_K$. This parameter describes the degree of aerodynamic coupling of the solid particles to the gas. Recall from the above discussion that $q_{\mathrm{fric}} \ll 1$ for particles that are tightly coupled to the gas and $q_{\mathrm{fric}} \gg 1$ for those that are only poorly coupled. One may view $q_{\mathrm{fric}}$ as a dimensionless stopping time. Particles that initially move at a velocity that differs from that of the gas are 'stopped' (i.e. become part of the bulk gas motion) on a timescale $q_{\mathrm{fric}}$ in units of the orbital time $\Omega^{-1}$.

Figure 10.5 shows the radial drift velocity as a function of $q_{\mathrm{fric}}$ for parameters that are approximately appropriate for that of a typical proto-planetary disk at a radius of 5 AU. The drift velocity peaks at $q_{\mathrm{fric}} \simeq 1$ at a value $v_r = \eta \, v_K/2$, that depends only upon the pressure gradient in the disk via the dependence of $\eta$ on the sound speed and surface density gradients. In the case shown in the figure the maximum radial drift velocity is $50 \, \mathrm{m \, s^{-1}}$. The particle size that corresponds to $q_{\mathrm{fric}}$, i.e. the ones that drift in the fastest, can be computed based on the formulae for the friction time in the appropriate drag regime (i.e. Epstein or Stokes). In the Epstein regime, for example, one finds from Eq. (10.8) a size

$$s(q_{\mathrm{fric}} = 1) = \frac{\rho}{\rho_m} \frac{v_{\mathrm{th}}}{\Omega_K}. \qquad (10.24)$$

At 5 AU in a disk with $\Sigma = 10^2 \, \mathrm{g \, cm^{-2}}$ and $h/r = 0.05$ the fastest drift occurs for a particle size $s \simeq 20 \, \mathrm{cm}$. In the inner disk (1–10 AU) the most rapid radial drift is realized for particle sizes in the 10 cm to a few m range. If we introduce a minimum radial drift time $\tau_{\mathrm{drift}}$, i.e. the

---

[3]See Armitage (2010), *Astrophysics of Planet Formation*, section 4.3, p. 119, but note differences in the adopted notation.
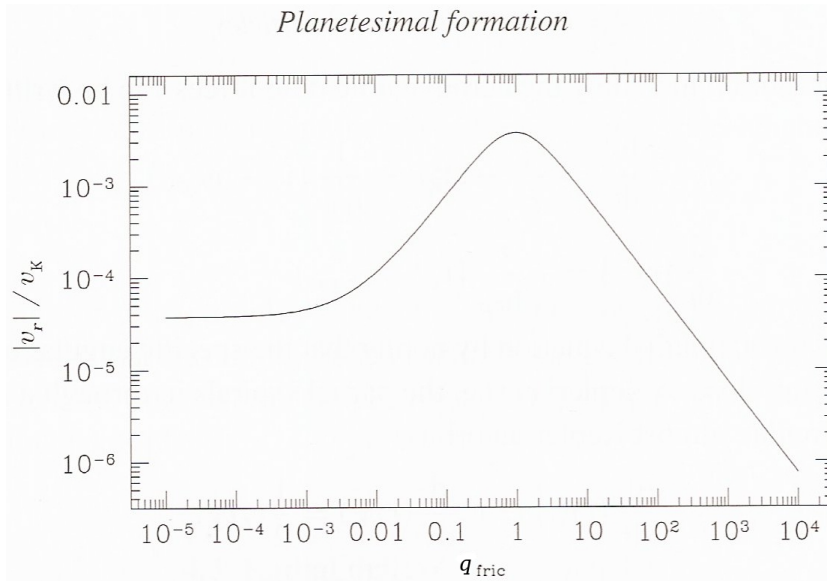
Figure 10.5: *The radial drift velocity of particles at the mid-plane of the proto-planetary disk is plotted as a function of the dimensionless stopping time $q_{\mathrm{fric}}$. It is assumed that $\eta = 0.0075$ and that $v_{r,\mathrm{gas}}/v_{\mathrm{K}} = -3.75 \times 10^{-5}$. These values are approximately appropriate for a disk with $h/r = 0.05$ and $\alpha = 0.01$ at 5 AU. The maximum radial drift velocity at $q_{\mathrm{fric}} = 1$ is $50\,\mathrm{m\,s^{-1}}$. From: Armitage 2010, Physics of planet formation.*

timescale evaluated at the peak of the curve in Fig. 10.5, for particles to drift to the central star

$$\tau_{\mathrm{drift}} = \frac{r}{|v_{r,\mathrm{peak}}|} \qquad (10.25)$$

one finds that throughout the planet forming disk this timescale is of the order of $10^3$ yr or less in the inner disk and $10^4$ yr or less in the outer disk for reasonable disk parameters. This is extremely short and is an important result, from which three conclusions can be drawn:

- ○ Planetesimal formation must be rapid, at least if it occurs through a cascade of pairwise collisions that lead to growth. If growth through the cm-m size scale were not very rapid the vast majority of the solid material in the disk would drift toward the star to be evaporated in the hot inner regions of the proto-planetary disk.

- ○ Radial redistribution of solids is very likely to occur. Radial flow of solid particles on a timescale shorter than the disk lifetime occurs not just at the peak of the radial velocity drift curve, but also for substantially smaller and larger particles.

- ○ Particle concentration may occur near local pressure maxima. In a negative gas pressure gradient regime, sizable solid particles experience a headwind so they drift inward, while gas in a regime of positive gas pressure gradient orbits faster than Keplerian,

yielding a tailwind that raises particles' orbits. At the interface of these regimes, near a local pressure maximum, particles are collected.

## Radial drift with aggregation

The size dependence of the radial drift velocity introduces a relative velocity between particles of different size, which may lead to growth via aggregation. Using similar arguments as that employed to describe aggregation during vertical settling, we may write – in the limit where all collisions are adhesive – for the growth rate of a particle of radius $s$ that collides primarily with very small particles that themselves drift in with negligible radial velocity

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \pi\, s^2\, |v_r|\, f_{\mathrm{dg}}\, \rho(z), \tag{10.26}$$

where the parameters other than $|v_r|$ have the same meaning as in Eq. (10.12). Focussing on the mid-plane, where the density is given by Eq. (9.27), we may determine the maximum size of the particles for which the growth timescale $\tau_{\mathrm{grow}} = m/(dm/dt)$ is shorter than the drift timescale $\tau_{\mathrm{drift}}$. This yields

$$s \lesssim \frac{3 f_{\mathrm{dg}}}{4\sqrt{2\pi}} \left(\frac{h}{r}\right)^{-1} \frac{\Sigma}{\rho_{\mathrm{m}}}. \tag{10.27}$$

Assuming that modest settling has already taken place such that $f_{\mathrm{dg}} \sim 0.1$, and taking other appropriate values $\Sigma = 10^3\,\mathrm{g\,cm^{-2}}$, $\rho_{\mathrm{m}} = 3\,\mathrm{g\,cm^{-1}}$, and $h/r = 0.05$, we find that particles with a size up to $s \simeq 2\,\mathrm{m}$ would collide with at least their own mass of other particles during their inward drift.

This is insufficient to alleviate the radial-drift barrier problem. Moreover, what turns out to be yet another serious problem is that the relative collision velocities between large rock-sized particles will very likely invalidate the assumption that collisions will lead to growth. Indeed, at the peak of the radial velocity drift curve it is likely that collisions will actually break up particles, i.e cause fragmentation. The inability of grains of rock-sizes to grow further in mass because of too high impact velocities is referred to as the *fragmentation barrier*. We will discuss particle-particle interaction is more detail in the next chapter.

**Exercise 10.1**

a) The definition of the settling time, Eq. (10.11), seems to ignore that $v_{settle}$ depends on $z$, even for a constant gas density $\rho$. Realizing that $v_{settle} = -dz/dt$, how would you define the characteristic settling time for a constant gas density and in the limit $z \ll r$ (hence $r$ is essentially contant)?

b) Is the interpretation of your definition of the settling time the same as the definition of $\tau_{settle}$ in Eq. 10.11.

**Exercise 10.2**

Consider a small dust particle of sub-micron size in a proto-planetary disk. The particle is located about one scale height above the disk mid-plane, some 10 AU from the central star. Discuss at least three possible fates of the grain.

**Exercise 10.3**

Notice that in Fig. 10.4 the mass of a particle at arrival in the disk mid-plane while settling from a great vertical height and growing by aggregation appears independent from its initial radius.

a) Derive that

$$m_{final} = \frac{\pi}{384} \frac{(\Sigma f_{dg})^3}{\rho_{mp}^2} \tag{10.28}$$

assuming the particle settles from an infinitely high starting position above the disk mid-plane and has negligible initial mass.

b) Compute this mass in gram for the parameters given in the caption of Fig. 10.4

c) Does every initial condition lead to a mass $m_{final}$ for the grain on arrival in the disk mid-plane?

**Exercise 10.4**

Write a computer program that solves the linear equations Eqs. (10.12) and (10.14) for dust grains of initial size $s = 1, 100, 1000\,\mu$m settling to the disk mid-plane from an initial height of $z = 5h$, where the scale height $h = 0.02$ AU. Use other relevant properties as given in the caption of Fig. 10.4. So, recover the result for $1\,\mu$m grains shown in Fig. 10.4 and explore the settling of grains that initially are much larger.

**Exercise 10.5**

In a turbulent disk, small dust grains suspended high above the disk mid-plane are prevented from settling because of the continuous stirring of the gas in turbulent swirls and eddies. They first need to experience substantial *in situ* growth before settling may happen. In this exercise, we are concerned with calculating the time scale required to grow a particle at a distance $z$ above the disk mid-plane from an initial radius $s_\circ = 0.1\,\mu$m to a size of 1 mm, where turbulence can no longer keep the particle suspended and settling commences. The material density $\rho_\mathrm{m}$ of the particles is $1.6\,\mathrm{gr\,cm}^{-3}$.

At 1 AU from the star the gas mid-plane density $\rho_\mathrm{mp} = 6 \times 10^{-10}\,\mathrm{gr\,cm}^{-3}$, the scale height $h = 0.02\,$AU, and the thermal gas velocity $v_\mathrm{th} = 10^5\,\mathrm{gr\,cm}^{-3}$. The mean molecular weight of the gas particles is $\mu = 2.3$. The dust-to-gas ratio $f_\mathrm{dg} = 0.01$.

We assume that the velocity of the small particles suspended in the gas high above the disk mid-plane is the Brownian velocity $v_\mathrm{Br}$. We further assume that only the one particle that we consider is growing in size, and hence that it remains embedded in a sea of particles that all have size $s_\circ$. As our particle will quickly grow much larger, the relative velocity of the small particles and our particles is well described by the constant value $v_\mathrm{Br}$.

Our particle then grows in mass at a rate (see also Eq. 10.12)

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \pi s^2\, v_\mathrm{Br}\, f_\mathrm{dg}\, \rho(z), \tag{10.29}$$

where $\rho(z)$ may be derived from Eq. (9.25).

a) We assume that dust particles are in thermal equilibrium with the gas, i.e. that the kinetic energy of a gas particles is the same as that of a dust grain. Compute the Brownian velocity $v_\mathrm{Br}$ of the dust particles of size $s_\circ$.

a) Derive a differential equation for the growth rate $ds/dt$ of our particle.

b) Solve this differential equation. At $t = 0$, $s(t = 0) = s_\circ$.

c) Compute the *in situ* growth time to $s = 1\,$mm at $z = 2h$ in year.

d) At $z = 5h$ the growth time of the small particles is so large that they will never reach sizes large enough to settle toward the disk mid-plane. What, in the end, will happen to these particles?

# 11

# Building planets (part II) - the intermediate regime

The most poorly understood phase in planet formation is the one in which particles grow from centimeter-sized aggregates to kilometer-sized planetesimals. We first discuss the possibility to achieve this through a continued growth by aggregation, even though in the previous chapter we identified two fundamental obstacles for this to happen: the *fragmentation barrier* and the *radial-drift barrier*. Next, we discuss particle concentration leading to the direct formation of planetesimals by gravitational collapse of the densest particle agglomerations. Last, we discuss gravitational instability of the solid state component of the disk, also resulting in direct planetesimal formation.

## 11.1 Planetesimal formation by sweep-up

In the previous chapter, we assumed that a large particle settling in the mid-plane or drifting towards the central star sweeps up all small particles in its path. Laboratory experiments show that the outcome of an interaction of two dust grains depends on a number of factors, including the target to projectile mass ratio, the compactness of the aggregates (compact or fluffy; see Fig. 11.1), the spin rate of the particles and the relative orientation of their rotation axes, and impact angle. Of key importance is the impact speed. Roughly speaking, experimentally established outcomes can be divided in five main categories (Güttler et al. 2010; see also Fig. 11.2). In order of increasing relative impact velocity these are:

- Sticking. The two particles stick together on impact. Mass is a conserved quantity, i.e. the final mass is the sum of the masses of the two components.

- Bouncing. Typically rather compact particles may simply bounce off at impact.

- Mass transfer. Mass is not conserved in the interaction, a modest spray of small debris
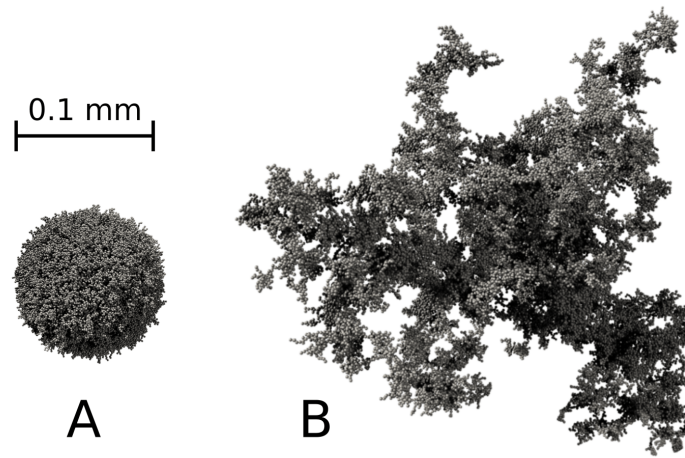
Figure 11.1: *Two dust aggregates grown in computer simulations. Aggregate A has grown by many hit-and-stick collisions with micrometer-size grains. Aggregate B has grown by colliding with particles similar to itself, which is closer to what is expected to occur in proto-planetary disks. From: Seizinger et al. 2013.*

particles is created. However, the main post-impact component is more massive than the most massive pre-impact component.

- Erosion. Mass is not conserved in the interaction, a fairly large spray of small debris is created such that the main post-impact body is less massive than the most massive pre-impact component.

- Destruction or fragmentation. The impact velocity is so high the two bodies shatter at impact, creating a swarm of smaller particles.

For the smallest particles the role of surface forces is of key importance. Dust grains that are non-magnetic and electrically neutral, adhere as a consequence of induced dielectric forces if the grains collide gently enough. For a $0.5\,\mu$m sized spherical grain sticking may occur if $\Delta v \lesssim 1-2\,\mathrm{m\,s}^{-1}$. For similar-sized grains of irregular shape there is a non-zero probability of adhesion up to collision velocities of as large as $100\,\mathrm{m\,s}^{-1}$. For larger particles, the declining ratio of surface area to mass means that surface forces become less important. In this case the collision outcomes are controlled increasingly by the ability of particles to dissipate energy upon collision. Solid bodies fail to dissipate sufficient energy and tend to rebound rather than stick at typical collisions velocities. It is reasonable to assume that the typical structure of cm-sized and larger particles may instead resemble a loosely bound aggregate of dusty or icy particles. Such aggregates can dissipate a greater fraction of the kinetic energy of impact internally, and experiments confirm that this can result in sticking probabilities that are substantially more favourable for ongoing particle growth. Growth of such aggregate particles via collision is plausible for $\Delta v \lesssim 1\,\mathrm{m\,s}^{-1}$. Quite violent impacts with $\Delta v \gtrsim 10\,\mathrm{m\,s}^{-1}$ may
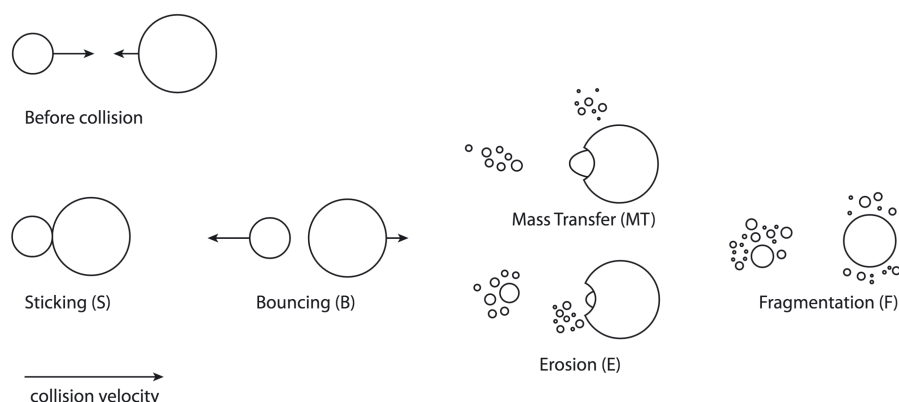
Figure 11.2: *A sketch of some of the most important possible outcomes of a collision between dust grains. From: Windmark et al. A&A 2012, 540, A73.*

still yield net growth, although this is accompanied by a spray of small debris (Wurm et al. 2005; Dominik et al. 2007).

Numerical models of the collision of dust particles aim to recover the collision experiments. Though all types of collisions are seen, one does find interesting discrepancies between the simulations and the laboratory experiments, where the simulations have difficulties reproducing the bouncing events and generally observe much higher fragmentation thresholds (Windmark et al. 2012). Movies showing theoretical results can be found on the YouTube channel of Alexander Seizinger[1]

Implementation of the properties of the experimental outcomes of collisions in dust-size evolution codes allows to study the grain-growth process in more detail. Okuzumi & Hirose (2011) developed a collision model where growth was possible for silicates for velocities up to $7 \, \text{m} \, \text{s}^{-1}$, and for ices up to $70 \, \text{m} \, \text{s}^{-1}$. By incorporating relative velocities in the dead zone extracted from MHD simulations, they were able to form planetesimals made of ice, but not of silicates. To illustrate some of the intricacies, in case of a large ratio between target and projectile mass, fragmentation may still lead to net growth. Once the projectile strikes the target, a sizable part of the spray of debris may stick to the surface of the target by van der Waals forces – a process called *fragmentation with mass transfer*. Studying this process, Wurm et al. (2005) and Teiser & Wurm (2009a) found that the accretion efficiency even increased with velocity, and could be as high as 50% of the mass of the fragmented projectile. Teiser & Wurm (2009b) also studied multiple impacts over the same area, and could conclude that growth was possible even then, without the newly accreted material being eroded.

So, perhaps it is possible to grow past the fragmentation barrier after all. Windmark et al. (2012) perform dust-size evolution simulations for a physically motived scheme of collisional outcomes (see Fig. 11.3), taking into account the laboratory results of bouncing experiments

---

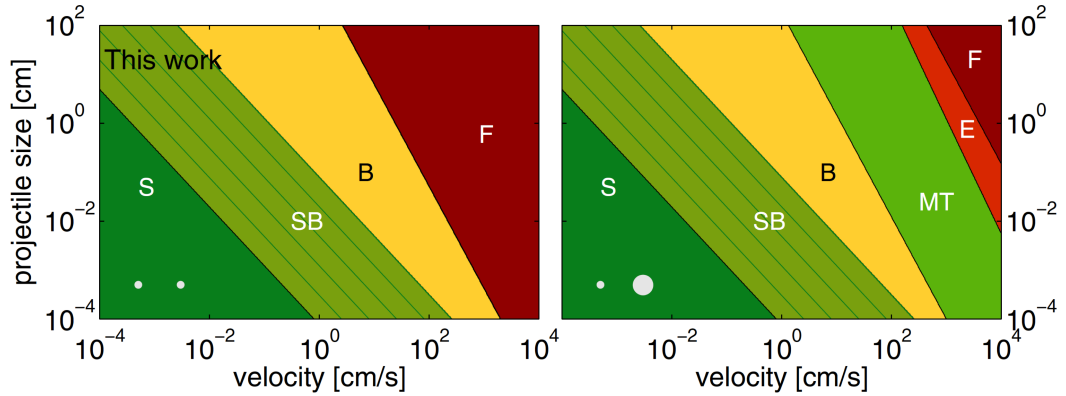[1]See https://www.youtube.com/user/aseizinger/featured

Figure 11.3: *Model describing the outcome of collisions for equal-sized collisions (left) and differently sized collisions (right). In the right panel, the target has a mass that is 1000 times the mass of the projectile. Green regions mark collisions that are growth-positive for the target, yellow marks growth-neutral, and red marks growth-negative. S marks sticking; SB marks the boundary zone between sticking and bouncing, with green parallel lines each marking a decrease in sticking probabilitly by 20% from left to right. B marks bouncing; MT net mass transfer; E net erosion, and F fragmentation. From: Windmark et al. A&A 2012, 540, A73.*

as well as the fragmentation-with-mass-transfer effect. For the general dust population, they find that bouncing collisions prevent any growth above millimeter-sizes. This is referred to as the *bouncing barrier* (Zsom et al. 2010). However, if a small number of cm-sized particles are introduced, for example by either vertical mixing or radial drift, they can act as a catalyst and start to sweep up the smaller particles. The bouncing barrier is here even beneficial, as it prevents the growth of too many large particles that would otherwise only fragment among each other, and creates a reservoir of small particles that can be swept up by larger bodies. Again, for this process to work, a few seeds of cm-sized or larger have to be introduced.

To investigate this scenario, we consider a single large body of mass $m$ and radius $s$ which moves around in a sea of small particles of mass $m_s$ and radius $s_s$ that can not grow further because they encounter the bouncing barrier. The growth of the large body can then be described by

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \sigma \, \Delta v \, \epsilon_{\mathrm{net}} \, n_s \, m_s \tag{11.1}$$

where $\sigma = \pi \left(s + s_s\right)^2 \simeq \pi s^2$ is the collisional cross-section, $\Delta v$ the relative velocity, $n_s$ the number density of the small particles, and $\epsilon_{\mathrm{net}}$ the average net mass-gain efficiency of a collision. The change in size $s$ can be written as $dm = 4\pi \rho_m \, s^2 \, ds$, where $\rho_m$ is the material density of the large body. Later on, we will adopt $\rho_m = 1.6 \, \mathrm{gr \, cm^{-3}}$. We can now write

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \frac{1}{4} \frac{\rho_s}{\rho_m} \, \epsilon_{\mathrm{net}} \, \Delta v, \tag{11.2}$$

where $\rho_s = n_s \, m_s$ is the mass density of the small particles.

Different processes may contribute to the relative velocity: Brownian motion, turbulence, and both azimuthal and radial drift. We assume here that $\Delta v$ is only caused by turbulence. It may be estimated by (Weidenschilling & Cuzzi 1993; Cuzzi et al. 2001)

$$\Delta v = \sqrt{\left(\frac{9}{2}\,\alpha\,\text{Stk}\right)}\cdot a, \tag{11.3}$$

where $\alpha$ is the Shakura & Sunyaev viscosity (Eq. 9.57), $a$ the sound speed in the gas (Eq. 3.3), and Stk the Stokes number for the boulder-sized particle. Below, we will adopt $\alpha = 10^{-3}$. For small particles the Stokes number, denoting how strongly a particle is coupled to the surrounding gas, can be written as

$$\text{Stk} = \frac{\rho_\text{m}\,s}{\Sigma}\frac{\pi}{2}, \tag{11.4}$$

where $\Sigma$ is the surface density of the gas (Windmark et al. 2012). We further assume $\rho_\text{s}$ to be constant and unaffected by the sweeping-up of the large boulder. Equation (11.2) can then be solved analytically. This yields

$$s(t) = \left(\frac{1}{8}\,\epsilon_\text{net}\frac{\rho_\text{s}}{\rho_\text{m}}\sqrt{\frac{9}{4}\frac{\rho_\text{m}}{\Sigma}\pi\alpha}\cdot a\cdot(t-t_\circ)+\sqrt{s_\circ}\right)^2, \tag{11.5}$$

where at time $t_\circ$ the size of the large object is $s_\circ$.

When the boulder-sized particle has grown large enough, it is hardly affected by the vertical mixing from turbulence, and is mostly found near the mid-plane. Assuming the gas density profile of the minimum-mass solar nebula (see Eq. 9.22), the gas surface density at 3 AU is $330\,\text{gr}\,\text{cm}^{-2}$. If we further assume that all dust mass can be found in the small particles and that for this population the dust-to-gas ratio is 0.01, we obtain a small particle surface density of $3.3\,\text{gr}\,\text{cm}^{-2}$. Following Windmark et al. (2012) we adopt $T = 115\,\text{K}$ at this location, hence a sound speed $a = 0.65\,\text{km}\,\text{s}^{-1}$ for a gas mixture with mean molecular weight $\mu = 2.3$. Using Eq. (9.26), we obtain a gas scale height $h = 1.7 \times 10^{12}\,\text{cm} = 0.1\,\text{AU}$. For a Gaussian vertical density profile (Eq. 9.27) we then get $\rho_\text{s} \sim 1 \times 10^{-12}\,\text{gr}\,\text{cm}^{-3}$, realizing that the turbulent motions cause the small particles to mix well with the gas. In the red curve of Fig. 11.4 the solution of the boulder-sized particles is shown assuming a net mass-transfer efficiency $\epsilon_\text{net} = 0.006$, which matches laboratory experiments for the impact of a 1 mm projectile body into a 1 cm target body by Teiser & Wurm (2009a). In the same figure, the growth solution for the relative velocities for each particle-particle pair calculated from all four velocity components and matching (model for the) net mass-transfer efficiency is given in black. In this full simulation, $\epsilon_\text{net}$ typically ranges from 0 (for projectiles at fairly low $\Delta v$ bouncing off of the target) to 0.15 (for fairly small projectiles hitting large targets at not too high $\Delta v$). Our simple model compares well to the growth rate of this full simulation for the first stage of growth. At later stages it differs greatly because of two reasons. First, our assumption of a constant $\epsilon_\text{net}$ means that at larger sizes the mass transfer is underestimated. Second, our relative velocity prescription Eq. (11.3) only considers turbulence. Moreover, it is only valid for small particles. To illustrate that it is indeed these deficiencies in our toy
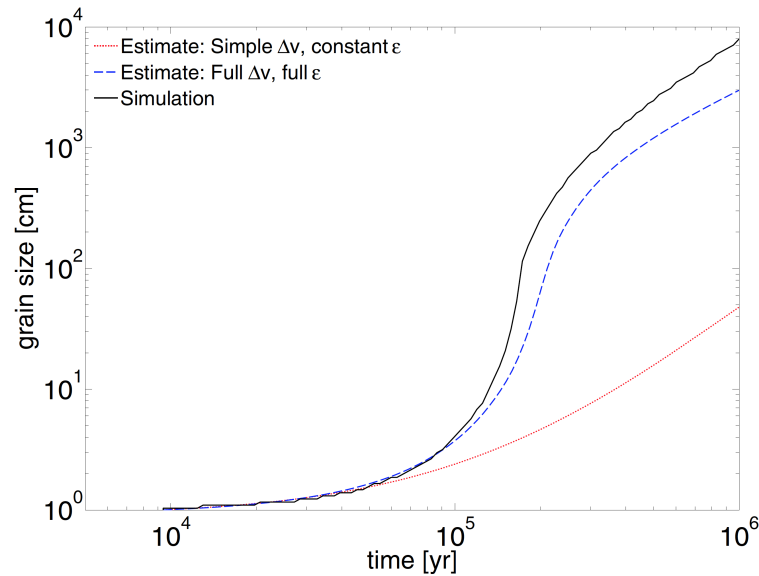
Figure 11.4: *The growth of a boulder-sized particle as in the simulations by Windmark et al. 2012 (black line). The red line corresponds to the simplified analytical estimate discussed in the text, and the blue line to an estimate using the full relative velocity and accretion efficiency prescriptions as adopted in the black line. From: Windmark et al. A&A 2012, 540, A73.*

model that explain the difference in later stages of growth, Fig. 11.4 also shows the solution of Eq. (11.2), taking into account all four velocity sources and the full net growth efficiency. The new solution, denoted with a blue dashed line, and the full simulation results now agree very well, only to differ slightly at the largest particle sizes. The enhanced growth is triggered by a population of intermediate-sized particles that form from fragmenting events between large bodies. The largest particles in the simulations are those that manage to grow by avoiding interactions with other large bodies. When they manage to interact beneficially with some of the intermediate-sized fragments, the growth rate increases steeply.

Still, in absolute terms the growth rate is relatively slow. The artificially inserted few 1 cm-sized seeds grow to ∼100 m in size on a timescale of 1 Myr. The reasons for the slow growth are mainly the low mass transfer-efficiency, the high turbulence that kicks the small particles away from the mid-plane where the boulder-sized objects are concentrated, and the low dust densities even at radii as small as 3 AU.

Though major hurdles remain for this planetesimal formation mechanism, the interesting aspect is that direct growth of planetesimals via dust collisions between particles of very different sizes seems to be a feasible route. Interestingly, the bouncing barrier introduced by Zsom et al. (2010) might be beneficial or even vital for the planetesimal formation, as it provides a natural way of ensuring that most of the dust population remain small. These small dust particles are ideal for the sweep-up process if larger bodies manage to form. Exactly how

these larger seeds would be introduced into the system needs to be explored in greater detail.

## 11.2 Planetesimal formation by particle concentration

As pointed out in the introduction to chapter 10 several ideas are being tested to overcome the radial-drift barrier. A leading theme in these ideas is that of *particle concentration* or *particle clumping* in the sedimented mid-plane layer. Local pressure maxima (pressure bumps), turbulent eddies, streams, spiral arms or dead zones may cause such concentration. Another relevant circumstance in which inward drift can be halted occurs in the presence of a massive planet, which tends to create an annular gap in the disk surface density near the location of its orbit. A disk with such a gap possesses a pressure maximum at the inner edge of the disk that is exterior to the planet. Solids with some range of sizes would be expected to accumulate at that location. Obviously, regarding this last possibility one has to ponder on the question as to how that first planet formed.

In any case, high local particle densities can lead to planetesimal formation via direct collisions or via gravitational collapse if the dust-to-gas ratio is roughly unity. Particle densities above the *Roche density* (see Exercise 11.1)

$$
\begin{aligned}
\rho_{\mathrm{R}} = \frac{9\Omega^2}{4\pi G} &= 4.3 \times 10^{-7} \left( \frac{d}{\mathrm{AU}} \right)^{-3} \quad \mathrm{gr\,cm^{-3}} \\
&= 316\,\rho \left( \frac{d}{\mathrm{AU}} \right)^{-1/4}
\end{aligned}
\tag{11.6}
$$

are bound against the tidal force from the central star and can contract towards solid densities. Here $d$ is distance to the central star. The latter equation relates the Roche density to the gas density $\rho$ assuming typical disk properties (see Exercise 11.2). It reveals that over-densities by a factor $\sim 10^2 - 10^3$ are required to initiate direct gravitational collapse.

A promising candidate mechanism for particle concentration are *streaming instabilities*. These instabilities are based on the notion that drag forces not only affect the motion of dust particles embedded in the gas, but that drag also produces a back reaction on the gas, increasing its velocity. Let us first recap the effects of drag on dust: The gas orbits at a somewhat slower velocity than dust as it profits from a pressure gradient that partially offsets the stellar gravitational pull. Solid particles are not supported by gas pressure and would orbit at Keplerian velocities in the absence of gas. The difference in velocities results in a headwind that causes the solid particles to spiral toward the central star as they loose momentum to aerodynamic drag. Now let us discuss the effect of drag on gas: The faster moving solids also produce a drag force on the gas, increasing its velocity, particularly so when the solids somewhat cluster in the gas. Such clustering reduces the headwind locally, allowing particles under the lee of the fast moving clusters to orbit faster and undergo less inward drift[2]. These relatively fast

---

[2]Of course, one need not explain streaming instabilities to Belgian students. It is a common phenomenon in the
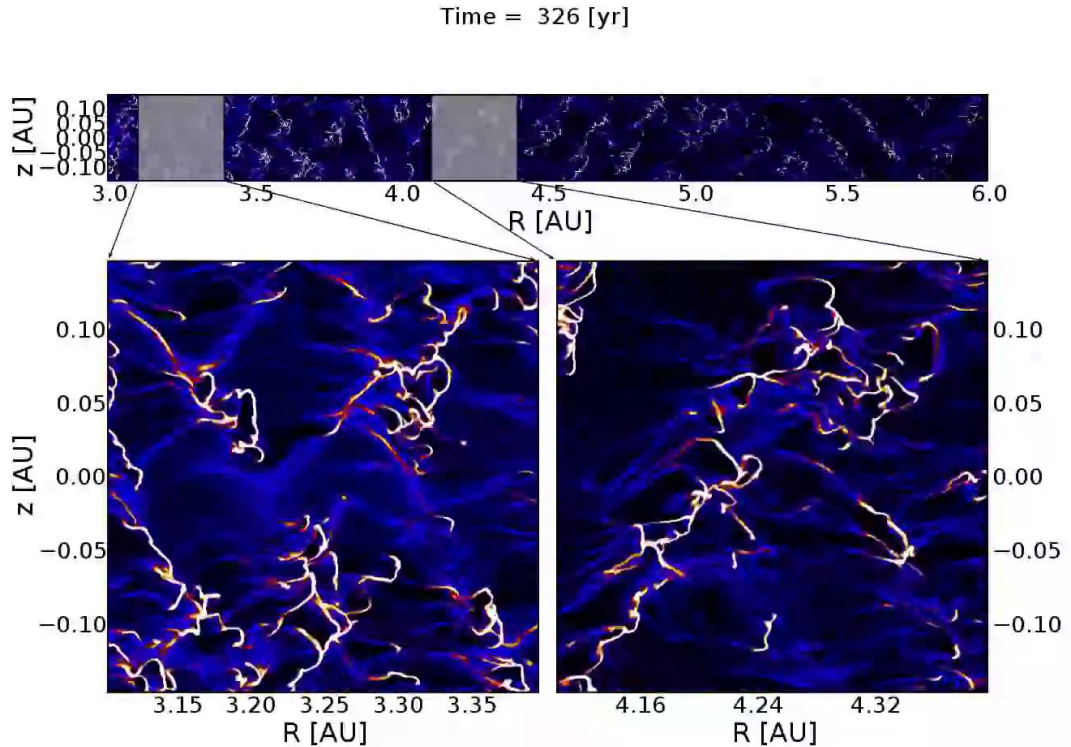
Figure 11.5: *2D simulation of two mutually interacting fluids, i.e. gas and dust, in a proto-planetary disk while neglecting the vertical component of gravity. The grains have a size of 50 cm and the dust-to-gas ratio is 1.0. The simulation lasts 3000 yr and shows a rapid growth of unstable modes in the dust component, which are consistent with modes obtained for that of the streamiing instability. The instability saturates when dust over-densities synchronize their velocity with the gas component. Despite the steady outward migration of the gas, dust blobs never really gain positive radial velocity and never cease to migrate inwards. From: Kowalik et al. 2013, MNRAS 434, 1460.*

moving particles overtake slower moving particles creating the lee zone, increasing the local density and further reducing radial drift. This fuels an exponential growth, the more so when clusters of particles start forming everywhere in the gas where conditions are favourable (i.e. where the solid to gas ratio $f_{\rm dg}$ is near or larger than unity) and relatively slowly (tangentially and radially) drifting clusters are overtaken by faster moving clusters.

In simulations such as shown in Fig. 11.5 the clusters form massive filaments that can grow or dissolve, and that can collide and merge or split into multiple filaments. Typical separations of filaments are $0.2h$. Their densities can exceed a thousand times the gas density, sufficient to

Tour of Flanders, when cyclists at the lead of the peloton create a lee zone for the cyclists behind them. These may easily overtake the leading racers, creating a concentration of cyclists, if these racers would not regularly apply their brakes to prevent this from happening.

trigger gravitational collapse and fragmentation of the filaments into bound clusters, leading to the formation of planetesimals the size of large asteroids ($\sim 100\,\mathrm{km}$) in as little as 25 years. The collapse of smaller ($1 - 10\,\mathrm{km}$) bodies is relatively slow and can extend over 1000 years because impact speeds are lower, avoiding the fragmentation and subsequent compacting of particles, resulting in the formation of *pebble pile planetesimals* with low densities. Collapsing swarms of dust particles with excess angular momentum (in the swarm) can fragment, forming *binary planetesimals*.

### Planetesimal formation via gravitational instability

A combination of vertical settling, radial drift, and photo-evaporation may result in regions in the disk where the mass in solid particles becomes comparable to the mass in gas. The solids then (locally) start to play a dynamical role. Gravitational instability within a dense layer of particles located close to the disk mid-plane might result in the prompt formation of planetesimals. The simplest version of this theory is known as the *Goldreich-Ward mechanism*. Although it has been shown to fail, closely related physical ideas remain important and this mode of planetesimal formation may be an option.

We do not discuss the stability of a razor-thin particle layer to gravitational collapse in great detail, but we do note that the physical considerations are identical to those discussed heuristically in Sect. 9.5 for the stability of a *gaseous* disk, with the particle velocity distribution $\sigma$ taking the place of the gas sound speed, and the stability criterium once again given by Toomre's $Q$ parameter (Eq. 9.63).

Assuming the particle layer can be treated as a fluid in which the two dimensional sound speed, defined in terms of the pressure $p$ and solid particle surface density $\Sigma_{\mathrm{solids}}$, $a = dp/d\Sigma_{\mathrm{solids}}$ is equivalent to the particle velocity dispersion $\sigma_{\mathrm{solids}}$, one may again describe the dynamics of the particle disk by the continuity and momentum equations, together with Poisson's equation for the gravitational field. A standard linear stability analysis results in a characteristic length scale $\ell$ for which the medium is most susceptible for instability. For smaller spatial scales pressure gradients stabilise disturbances; for larger spatial scales disturbances are smoothed out by differential rotation of the disk. In the intermediate regime self-gravity may dominate, causing instability. The typical length scale for this to occur is

$$\ell = \frac{2\,\sigma_{\mathrm{solids}}^2}{G\,\Sigma_{\mathrm{solids}}}. \tag{11.7}$$

The mass of clumps formed via gravitational instability is then of the order of

$$m_p \sim \pi\,\ell^2\,\Sigma_{\mathrm{solids}} \sim 4\pi^5 G^2\,Q_{\mathrm{solids}}^4\,\frac{\Sigma_{\mathrm{solids}}^3}{\Omega^4}, \tag{11.8}$$

where

$$Q_{\mathrm{solids}} = \frac{\sigma_{\mathrm{solids}}\,\Omega}{\pi G\,\Sigma} \tag{11.9}$$

is Toomre's parameter (see Eq. 9.63 for the equivalent expression in a gaseous disk). For instability to occur $Q_{\text{solids}}$ should drop to about or below unity. Adopting parameters that might be appropriate for the proto-planetary disk at 1 AU, i.e. $\Sigma_{\text{solids}} = 10\,\text{g}\,\text{cm}^{-2}$, the estimated planetesimal mass is $\sim 3 \times 10^{18}$ g. This corresponds to a spherical body with a radius of 5–10 km. If we assume – rather unrealistically – that once gravitational instability of the particle layer sets in collapse occurs on the free-fall timescale, then the formation time is very short, being less than a year for the parameters used here.

In principle a particle layer that fragments to form planetesimals could be composed of objects of any size, provided that the conditions for instability are met. The case that attracts the greatest interest, however, is that where the particles have sizes of the order of a cm or smaller. Such particles are small enough not to suffer the potentially devastating rapid radial drift of larger bodies, and there does not appear to be any particular obstacle to forming them rapidly via pairwise collisions. Subsequent instability of the layer made of small particles then has the feature – appealing to many – of rapidly forming planetesimals in a way that bypasses all of the potential hurdles involved in particle growth through the meter-scale regime.

To make the gravitational instability set in a very thin layer is required. For example, if the surface density in solids at 1 AU is $\Sigma_{\text{solids}} = 10\,\text{g}\,\text{cm}^{-2}$, we require a velocity dispersion of $\sigma = 10\,\text{cm}\,\text{s}^{-1}$ in order to attain $Q_{\text{solids}} = 1$. If the gas disk has $h/r = a/\Omega r = 0.05$ at this radius, the relative thickness of the particle layer $h_{\text{solids}}/r = \sigma_{\text{solids}}/\Omega r$ is $h_{\text{solids}}/h = \sigma_{\text{solids}}/a \sim 10^{-4}$. In the absence of gas, attaining such a razor-thin particle disk is not impossible – the vertical thickness of Saturn's rings, for example, is of the order of only 10 m. In the gas-rich environment of the proto-planetary disk turbulence may preclude the particle layer from ever becoming thin enough.

So, turbulent regions of the disk are unpromising sites for planetesimal formation via gravitational instability. Even if parts of the disk are not intrinsically turbulent, the thin disk (that may form) might make it turbulent: if the particles dominate the mass in the mid-plane, then the gas will be dragged along by the particles on a Keplerian orbit. Just above the mid-plane, the medium is deprived of dust and the gas is on a sub-Keplerian orbit. This vertical shear will be unstable to Kelvin-Helmholtz instabilities, thus inducing turbulent mixing and destroying the particle layer (Weidenschilling 1977).

Note that for instability to occur the system should fulfil the Toomre $Q_{\text{solids}}$ criterium (Eq. 11.9) ánd the Roche criterium (Eq. 11.6). The Toomre $Q_{\text{solids}}$ criteria provides a characteristic length scale, whereas the Roche criterium is scale free and only places requirements on the particle density. In Saturn's ring, for instance, the Toomre criterium is satisfied, but not the Roche limit (except in the outermost parts). So, in a sense, the particle concentration mechanism to form planetesimals discussed in Sect. 11.2 may be seen as a variant of the Goldreich-Ward mechanisms, in the sense that it describes gravo-turbulent planetesimal formation.
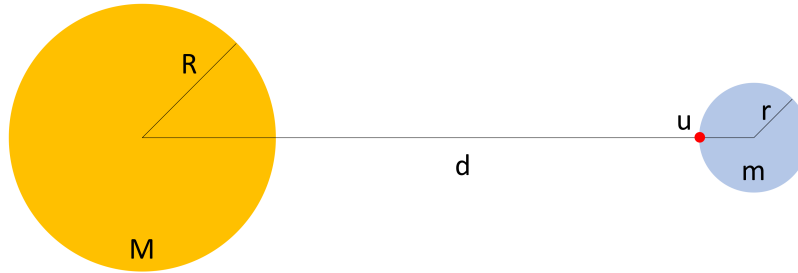
**Exercise 11.1**



Figure 11.6: *Geometrical situation for deriving the Roche radius and Roche density.*

We are concerned with deriving Eq. (11.6) for the Roche density. Consider the situation depicted in Fig. 11.6. A small test particle with mass $u$ is placed on the surface of a satellite that is closest to the primary star. The satellite has mass $m$ and radius $r$. The primary star has mass $M$ and radius $R$. We assume that $r \ll R$ and $R < d$, such that $r \ll d$. Several forces act on the test particle (see also Eq. 9.2):

$$
\begin{aligned}
F_{\text{gr}} &= \frac{G\,m\,u}{r^2} \\
F_{\text{tide}} &= \frac{2\,G\,M\,u\,r}{d^3} \\
F_{\text{cf}} &= \Omega_{\text{K}}^2\,u\,r = \frac{G\,M\,u\,r}{d^3}
\end{aligned}
$$

These are (1) the gravitation pull $F_{\text{gr}}$ on the mass $u$ by the satellite; (2) the tidal force $F_{\text{tide}}$ on the mass $u$ by the primary star, and (3) the centrifugal force from the orbital motion around the primary star. For the latter we assumed that the satellite is in synchronous rotation.

a) The tidal force is the difference in the primary's gravitational pull on the center of the satellite and on the edge of the satellite closest to the primary (where $u$ is positioned). Derive this equation.

b) Which forces balance each other?

c) Derive an equation for the *Roche radius* $r$ in terms of $m$, $M$, and $d$.

d) Use for the density of the satellite $\rho = 3m/4\pi r^3$ and derive Eq. (11.6).

**Exercise 11.2**

The goal of this exercise is to derive an equation expressing the Roche density $\rho_{\rm R}$ relative to the local gas density $\rho$, i.e. to derive the final term in Eq. (11.6). This tells us the overdensity $\rho_{\rm R}/\rho$ that needs to be achieved to induce direct gravitational collapse of particle concentrations.

We assume that the surface density of the proto-planetary is given by Eq. (**??**) and approximate the disk mid-plane temperature by

$$T_{\rm mp}(r) = 278 \left(\frac{r}{1\,{\rm AU}}\right)^{-1/2}.$$
(11.10)

The gradient of this temperature stratification is typical for a flaring disk, see Eq. (9.47). We assume vertical hydrostatic equilibrium in the disk, such that the gas scale height is given by Eq. (9.26) and the density in the mid-plane by Eq. (9.27). The mass of the central star is $1\,M_\odot$. We assume the temperature in the vertical direction to be constant (and given by $T_{\rm mp}$). The mean molecular weight $\mu = 2.3$. Other relevant relations are Eq. (3.3), specifying the sound speed, and Eq. (9.2), giving the orbital angular velocity or Kepler frequency.

a) Derive Eq. (11.6).

# 12

# Building planets (part III) - Terrestrial planets and gas giants

Once planetesimals of size 1-100 km have formed, gravitational interaction becomes the dominant mechanism through which particles interact and further growth is achieved. Though the physics involved is simple – it is Newtonian gravity only – the number of objects present will be of order $10^9 - 10^{11}$, i.e. too large to directly simulate their N-body evolution for long enough to have planets form. Therefore, for the earliest phases of terrestrial planet assembly a statistical approach is adopted. Once the number of dynamically significant bodies has reduced to some $10^3 - 10^4$ more or less moon-sized bodies, direct N-body simulations become necessary. Fortunately, for such numbers of objects such simulations are feasible.

## 12.1 Terrestrial planet formation

We discuss several aspects of the physics of collisions relevant for the formation of terrestrial planets from an ensemble of planetesimals.

### Gravitational focussing

For small particles the collision cross section of a particle of radius $s$ is simply the geometrical cross section $\pi s^2$. A massive object, however, will deflect the trajectory of other bodies and hence effectively has a cross section that is larger than its physical cross section – a phenomenon known as *gravitational focusing*.

Assume two planetesimal or planetary bodies that both have mass $m$ and radius $s$, initially moving towards each other with a relative velocity $\sigma$ on parallel trajectories as shown in Fig. 12.1. The impact parameter of these trajectories is $b$, i.e. the distance of closest approach in the absence of gravity. Gravity will deflect the orbit as shown in the figure, and they may pass by each other at a separation $r_c$ at closest approach, each then reaching a maximum
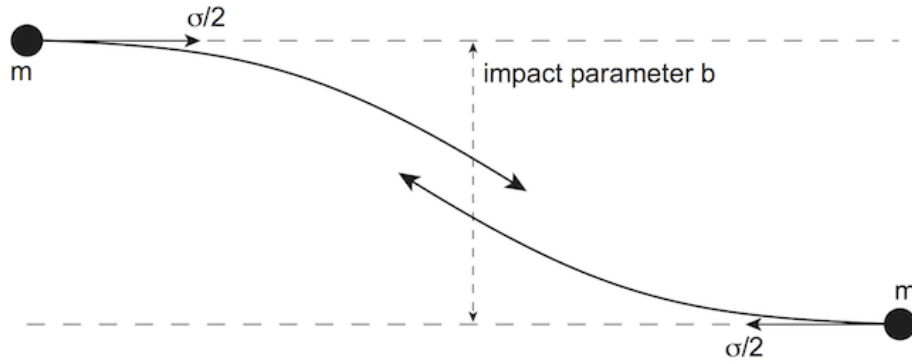
Figure 12.1: *Gravitational focussing effect of two particles of equal mass $m$ and relative velocity $\sigma$. From: Philip J. Armitage, Astrophysics of Planet Formation.*

velocity $v_{\max}$. We use conservation of energy and angular momentum to derive an expression for the effect of gravitational focussing on the cross section of these particles. Equating the total energy in the initial state of infinitely wide separation to the state of closest approach yields

$$\frac{1}{4}m\sigma^2 = mv_{\max}^2 - \frac{Gm^2}{r_c}. \tag{12.1}$$

Similarly, we may equate the total angular momentum Eq. 3.83 of the initial state to that of the state of closest approach. We find

$$v_{\max} = \frac{1}{2}\frac{b}{r_c}\sigma. \tag{12.2}$$

If $r_c < 2s$ the particles will collide. For larger $r_c$ they will miss each other and the encounter will end in a harmless flyby. Assigning the sum of the radii of the two bodies the parameter $r_s = 2s$, the square of the surface escape velocity of each of the particles $v_{esc}^2 = 2Gm/s = 4Gm/r_s$. Using this notation we find that the largest value of the impact parameter that will lead to a collision is

$$b^2 = r_s^2 + \frac{4Gmr_s}{\sigma^2} = r_s^2\left(1 + \frac{v_{esc}^2}{\sigma^2}\right). \tag{12.3}$$

The cross section is hence

$$A = \pi r_s^2\left(1 + \frac{v_{esc}^2}{\sigma^2}\right). \tag{12.4}$$

The first term is the geometrical cross section of the collision, the second term is the amplification factor to this cross section due to gravitational focusing. Gravitational focussing thus dominates collisions if the random velocity of the colliding bodies is small compared to the escape velocities from their surfaces. The gravitational focussing factor $\Theta = v_{esc}^2/\sigma^2$ is often referred to as the *Safronov number*.

## Hill sphere

The derivation of the effect of gravitational focussing assumes that the two interacting bodies are subject only to their mutual gravitational fields, and do not feel other forces. So, this ignores the gravitational force due to the central star that, if important, would mandate three-body dynamics to properly describe the situation. We may estimate the maximum radius around body $m$ within which its gravitational sphere of influence is more important than that of the more massive central star, of mass $M_\star$, by using a time-scale argument. This radius, referred to as the *Hill sphere radius*, is the distance $r$ from body $m$ for which a test particle in orbit around $m$ has the same orbital period $P = 2\pi\sqrt{r^3/Gm}$ as that of body $m$ at distance $d$ around the central star, $P = 2\pi\sqrt{d^3/GM_\star}$. We find

$$r_{\mathrm{H}} = \left(\frac{m}{M_\star}\right)^{1/3} d,$$

where $r_{\mathrm{H}}$ is the Hill sphere radius. A rigorous analysis, which is beyond the scope of these lecture notes, yields (see also Exercise 11.1)

$$r_{\mathrm{H}} = \left(\frac{m}{3M_\star}\right)^{1/3} (1 - \epsilon)\, d, \tag{12.5}$$

where $\epsilon$ is the eccentricity. So, for particles in eccentric orbits around a planet the Hill sphere radius is smaller than for those in circular orbits.

## Asymptotic velocity regimes of two-body interactions

There are two important asymptotic velocity regimes of two-body (read, for instance: planetesimal − large planetesimal or planetesimal − planet) interaction: *dispersion dominanted* and *Keplerian shear dominated*. The former is realized when the component of planetesimal velocity $\sigma$ is large compared to the *Hill velocity*,

$$v_{\mathrm{H}} = \sqrt{\frac{Gm}{r_{\mathrm{H}}}}, \tag{12.6}$$

i.e., the orbital velocity (see Eq. 9.1) of a test particle around the body of mass $m$ that has Hill radius $r_{\mathrm{H}}$. Think of the velocity dispersion (as set by the eccentricity- and inclination-distributions) of a swarm of small bodies in which the planet is embedded as being large, i.e. $\sigma \gg v_{\mathrm{H}}$. In this limit, the way in which such particles enter the Hill sphere and interact is determined by two-body dynamics. In the opposite regime, $\sigma \ll v_{\mathrm{H}}$, the relative speed of the bodies is mainly determined by Keplerian shear, i.e. differential rotation speeds. In this regime of small relative velocities, three-body effects must be considered.

## Trajectories of small bodies that move in the vicinity of planetary bodies

The Hill sphere radius can be used to compute the trajectories of small particles in the vicinity of a planetary body. The resulting trajectories are shown in Fig. 12.2. The situation that is
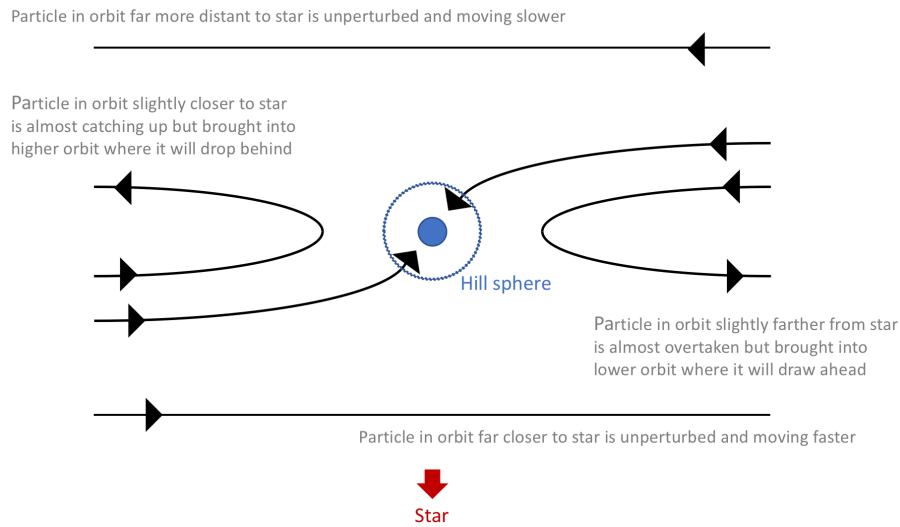
Particle in orbit far more distant to star is unperturbed and moving slower

Particle in orbit slightly closer to star is almost catching up but brought into higher orbit where it will drop behind

Hill sphere

Particle in orbit slightly farther from star is almost overtaken but brought into lower orbit where it will draw ahead

Particle in orbit far closer to star is unperturbed and moving faster

Star

Figure 12.2: *Trajectories of test particles in circular orbit around the star close to the orbit of a planetary body, as seen in the frame co-rotating with the planet. The planet moves to the right in the observers frame. Particles on orbits close to that of the planetary body never enter the latter's Hill sphere.*

sketched is for that of particles and planetary body in the disk mid-plane. Particles on circular orbits that are more than a few $r_H$ from the planetary body are essentially unperturbed by the presence of the planetary body. Particles within a few $r_H$, *but not too close in radial distance of the planetary body* enter the Hill sphere and may collide. Particles on almost the same orbit as the planetary body have a very small tangential velocity difference with the planet, so approach slowly, and actually attain a radial velocity large enough to be deflected to an orbit that causes it to again move away from the planet, either because it moves from an inner to an outer orbit where it starts to trail behind, or from an outer to an inner orbit where it starts to draw ahead of the planet.

## Possible outcomes of the collision of two planetesimal bodies

In the planetesimal range, the outcomes of collision of an impactor (the projectile) and a main body (the target) can be roughly divided into three categories:

- Accretion. The target body remains solid. All or most of the mass of the projectile becomes part of the mass of the target body. Small fragments may be ejected, but overall there is net growth, i.e. mass accretion.

- Shattering. The impact breaks up the target body into pieces, but the bulk of these pieces remain part of a single entity after gravitational re-accumulation (see Fig. 12.3). The structure of the shattered object resembles that of a rubble pile.

- Dispersal or Catastrophic Disruption. The impact is catastrophic and fragments the target into two, but likely many, pieces that do not remain gravitationally bound.

It is unlikely that the collision product will be a solid body if the initial target itself is a rubble pile. In this case, the collision could either disperse the pieces (the dispersal scenario) or merely rearrange them into a larger but still shattered object (the shattering scenario).

Of course, the outcome of a collision critically depends on several factors, including the target to projectile mass ratio, the impact parameter (i.e. the distance of closest approach if the cores of the two particles wouldn't feel each other), the composition and compactness of the bodies involved as well as their shape and rotation rate. A key factor is the impact velocity $v$. We may describe the outcome of a collision somewhat more quantitatively by introducing the *specific energy of the impact*, i.e., the kinetic energy of the projectile body per gram of the target body

$$Q \equiv \frac{mv^2}{2M}, \tag{12.7}$$

where $m$ is the projectile mass and $M$ the target-body mass. We may then define the three categories in the following way: $Q < Q_S^*$, where $Q_S^*$ is the threshold specific energy of impact for shattering; $Q_S^* < Q < Q_D^*$, where $Q_D^*$ is the threshold specific energy of impact for dispersal, and $Q_D^* < Q$. Benz & Asphang (1999) fit computational results for the catastrophic disruption threshold to the functional form

$$Q_D^* = Q_0 \left(\frac{R}{1\,\text{cm}}\right)^a + B\,\rho\,\left(\frac{R}{1\,\text{cm}}\right)^b, \tag{12.8}$$

where $R$ is the radius and $\rho$ the material density (in $\text{g}\,\text{cm}^{-3}$) of the target body, and $Q_0$, $B$, $a$, and $b$ constants. This functional form shows that $Q_D^*$ is described by two distinct physical regimes, represented by the two terms. The first term on the right describes the regime of relatively small targets (up to $\sim 100 - 1000\,\text{m}$), where the material strength of the target controls the outcome of the impact. In this *material strength dominated regime* $a < 0$, which implies that the material strength of bodies declines with increasing size (owing to the greater prevalence of defects that lead to cracks). In the *self-gravity dominated regime* $b > 0$. Large bodies are held together primarily by gravitational forces. Larger bodies will have a larger (specific) gravitational potential energy; see Eq. 3.55, from which it follows that this energy is proportional to $\rho R^2$). This motivates that $Q_D^*$ is increasing with size.

Table (12.1) provides the material density $\rho$ and relevant constants for a number of materials. For small target particles these rely on laboratory experiments and for large target particles on numerical N-body simulations. These numbers should be considered rough estimates as catastrophic disruption depends sensitively on the make-up of the target and exact impact properties and it would be unwise to place too much trust in precise numbers. Still, to get a rough feeling of numbers: the weakest icy bodies have minimum $Q_D^* \sim 10^5\,\text{erg}\,\text{gr}^{-1}$, while the strongest planetesimals have minimum $Q_D^* > 10^6\,\text{erg}\,\text{gr}^{-1}$. Figure 12.4 shows graphical representations of the dependence of $Q_D^*$ on target radius. Most of the results given in Table 12.1 are shown as well.
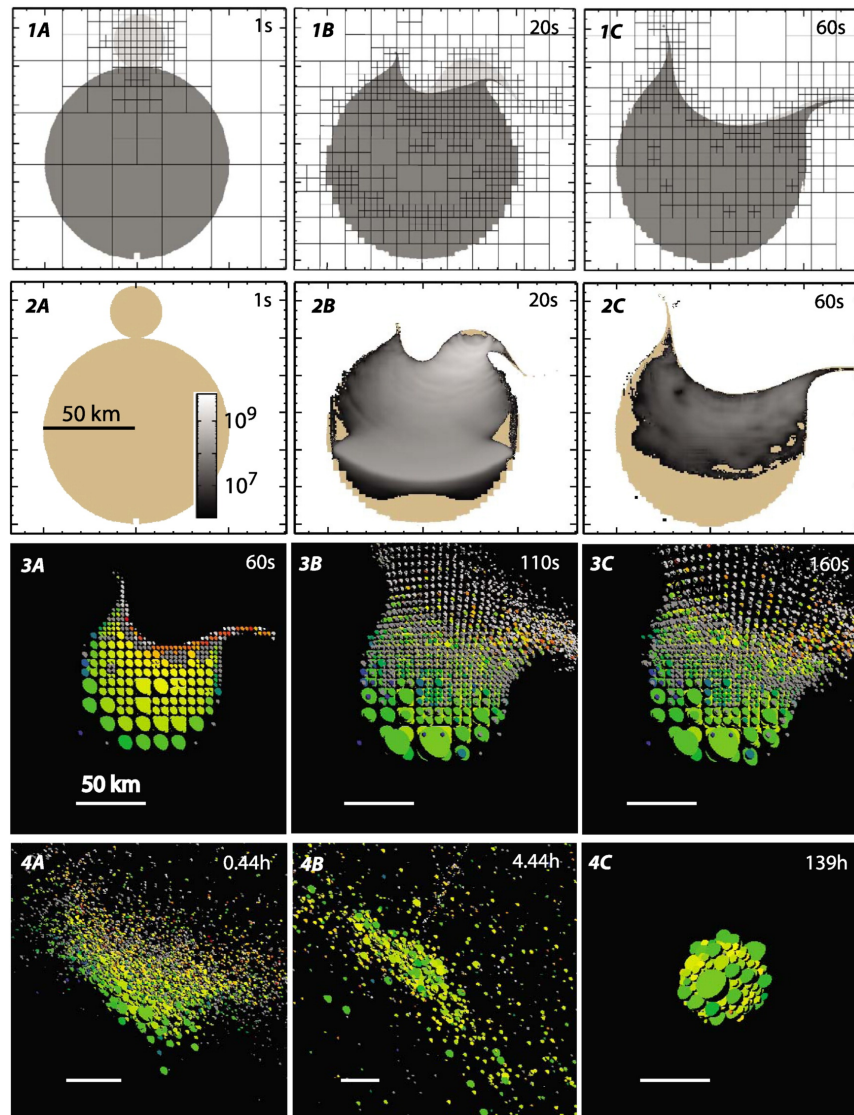
Figure 12.3: *3D simulation of the impact and gravitational re-accumulation phase of the collision of two basalt spheres of radius 14 km (projectile) and 50 km (target), using a Eulerian shock physics code for the initial phase (first 60 s) and an N-body gravity code for the later phases. The impact speed is* $1.8\ \mathrm{km\,s^{-1}}$; *the impact angle is 45 degrees. Top row: Projectile (light grey), target (dark grey), and the adaptive mesh. Row 2: The projectile and target (beige) and the pressure due to the impact (grey scale in Pa). Rows 3 and 4: 3A is a cross-section ($y = 0$ plane) of the target at 60s, after handoff from the Eurerian shock code. Colours represent the peak pressure attained during the impact (logarithmic range of 0.01 to 7 GPa). 3B-4C show the entire 3D object. The scale bar is 50 km in all frames. The last frame shows only the largest re-accumulated, post-collision remnant which equilibrates to 45% of the target mass in this simulation. From: Leinhardt & Stewart, 2009, Icarus 199, Issue 2, 542.*

| Material | Material strength | $\rho$ gr cm$^{-3}$ | $v$ km s$^{-1}$ | $Q_0$ erg gr$^{-1}$ | $B$ erg cm$^3$ gr$^{-2}$ | $a$ | $b$ |
|---|---|---|---|---|---|---|---|
| H$_2$O-ice | weak | 0.930 | 1.0 | $1.3 \times 10^6$ | 0.09 | -0.40 | 1.30 |
| H$_2$O-ice | strong | 0.917 | 0.5 | $7.0 \times 10^7$ | 2.1 | -0.45 | 1.19 |
| H$_2$O-ice | strong | 0.917 | 3.0 | $1.6 \times 10^7$ | 1.2 | -0.39 | 1.26 |
| Basalt | strong | 2.7 | 3.0 | $3.5 \times 10^7$ | 0.3 | -0.38 | 1.36 |
| Basalt | strong | 2.7 | 5.0 | $9.0 \times 10^7$ | 0.5 | -0.36 | 1.36 |

Table 12.1: *Catastrophic disruption threshold fitting formula, which describes how $Q_D^*$ depends on compactness (weak or strong) size of the target body and impact velocity $v$. From: Benz & Asphaug (1999; Icarus 142, 5) and Leinhardt & Stewart (2009; Icarus 199, 542 ) using experimental results in the strength dominated regime and numerical hydrodynamics simulations of collisions in the gravity dominated regime.*

### Runaway growth

Let us assume that a relatively massive body of mass $m$, radius $s$, and surface escape velocity $v_{\text{esc}}$ is embedded within a swarm of smaller planetesimals. The planetesimal swarm has a local surface density $\Sigma_{\text{p}}$ and an isotropic velocity dispersion $\sigma$. The planetesimal disk has half-thickness $h = \sigma/\Omega$ (see Eq. 9.26) and the volume density of planetesimals is $\rho_{\text{p}} = \Sigma_{\text{p}}/h = \Sigma\Omega/\sigma$. We will assume that the planetesimal collisions result in accretion, but note that in reality many collisions have sufficient energy to catastrophically disrupt the bodies. Current calculations (e.g., Stewart & Leinhardt 2009) suggest that gravity 'wins' only for planetesimals of size $s \overset{>}{\sim} 10\,\text{km}$. This aside, within our set of assumptions the growth rate due to collisions is given by

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \rho_{\text{p}}\,\sigma\,\pi s^2\,(1+\Theta) = \Sigma_{\text{p}}\,\Omega\,\pi s^2\,(1+\Theta) \tag{12.9}$$

where $\Theta = v_{\text{esc}}^2/\sigma^2$ is the Safronov number.

We may derive simple solutions for two special cases. First, we assume that $(i)$ the Safronov number is a constant, and $(ii)$ that the surface density of the planetesimal swarm does not change with time. Using $dm = 4\pi\,s^2\rho_{\text{m}}\,ds$, where $\rho_{\text{m}}$ is the density of the proto-planet, we find for the growth of the proto-planets's radius

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \frac{\Sigma_{\text{p}}\,\Omega}{4\rho_{\text{m}}}\,(1+\Theta). \tag{12.10}$$

As the right-hand side is constant, we find that the radius grows linearly in time, i.e. $s \propto t$. In the limit of negligible graviational focussing ($\Theta \ll 1$) we find for a solar mass star

$$\frac{\mathrm{d}s}{\mathrm{d}t} = 15.7\left(\frac{\Sigma_{\text{p}}}{10\,\text{gr cm}^{-2}}\right)\left(\frac{\rho_{\text{m}}}{1\,\text{gr cm}^{-3}}\right)^{-1}\left(\frac{d}{1\,\text{AU}}\right)^{-3/2}\quad\text{cm yr}^{-1}, \tag{12.11}$$
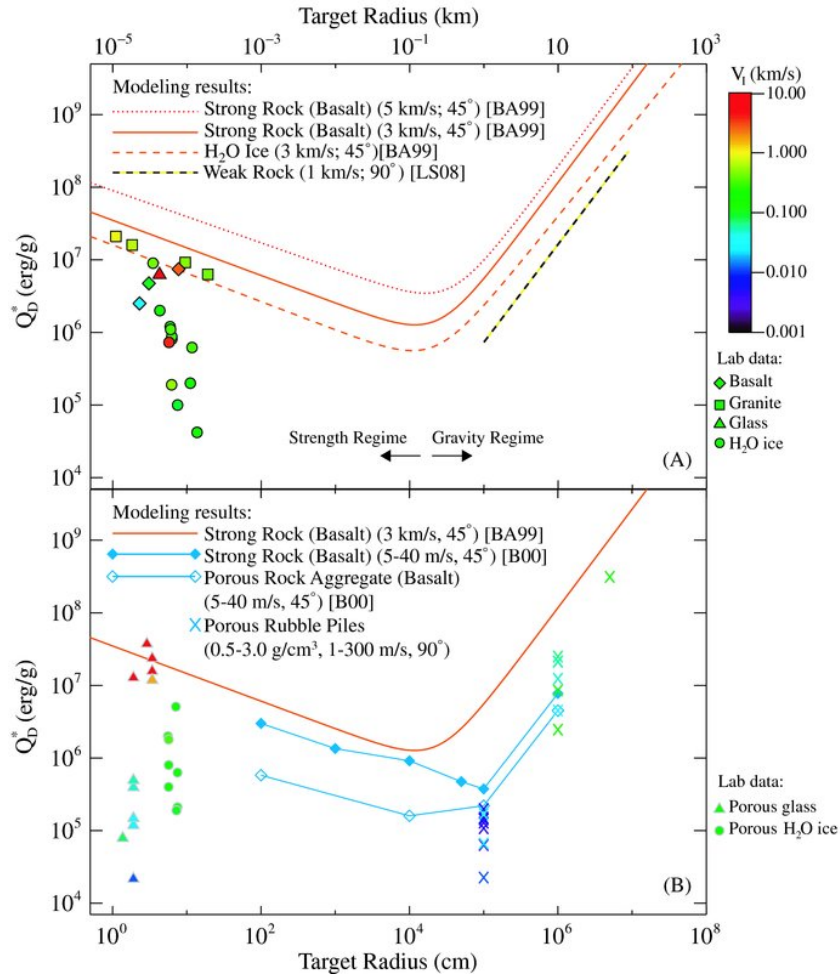
Figure 12.4: *The specific energy $Q_D^*$ for catastrophic disruption of solid bodies is plotted as function of the target body's radius. Results given in Table 12.1 are shown as well: BA99 is Benz & Asphaug 1999; LS08 is Leinhardt & Stewart 2009. Color denotes impact velocity. (A) Nonporous materials (strong): numerical results (Benz & Asphaug 1999; Leinhardt & Stewart 2009) and laboratory experiments. (B) Porous materials (weak): laboratory data on glass and ice and numerical experiments on aggregate rock (Benz 2000) and rubble piles. High- and low-velocity values for nonporous basalt are shown as reference. From: Stewart & Leinhardt, 2009, ApJL 691, 2.*

where $d$ is the distance to the central star. To grow a solid icy body of $\rho_m = 1 \, \mathrm{gr \, cm^{-3}}$ up to a size of 100 km at the location of Jupiter, where $\Sigma_p \simeq 10 \, \mathrm{gr \, cm^{-3}}$ in the proto-solar disk, would take 7.5 Myr. Given that the typical lifetime of the gas-rich phase of proto-planetary disks is a few Myr (see Fig. 9.13), we immediately conclude that to build bodies in the outer solar system large enough to accrete gas and to become gas giants requires strong gravitational focusing ($\Theta \gg 1$).

Given the necessity of gravitational focussing, let us consider the limiting case $\Theta \gg 1$. In this

regime Eq. (12.9) reduces to

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \Sigma_{\mathrm{p}}\,\Omega\,\pi s^2 \Theta = \frac{2G\,\Sigma_{\mathrm{p}}\,\Omega}{\sigma^2}\,s\,m = 2\left(\frac{3}{4\pi}\right)^{1/3}\frac{G\,\Sigma_{\mathrm{p}}\,\Omega}{\sigma^2\,\rho_{\mathrm{m}}^{1/3}}\,m^{4/3} \propto m^{4/3}. \tag{12.12}$$

Recall that this result has been obtained by assuming that the properties of the planetesimal swarm are fixed (i.e. the growing proto-planet neither excites the velocity dispersion nor consumes a significant fraction of the planetesimals). So, the growth rate increases for increasing $m$: an initially more massive body grows faster than its less massive cousin. This is called *runaway growth*, and demonstrates that if gravitational focussing dominates then the growth of planetesimals can be very rapid. In reality, this rapid growth is limited due to three reasons: (1) In a real system the number of close encounters exceeds the number of physical collisions, and repeated close encounters tend to increase the velocity dispersion $\sigma$, limiting the efficiency of gravitational focussing $\Theta$. (2) The collision velocities can be large enough to cause catastrophic disruptions, and (3) even if runaway growth does occur, it can only proceed until the local mass reservoir has been exhausted.

Of course, runaway growth can only proceed until the local mass reservoir has been exhausted. Planetesimals can only accrete from within their gravitational region of influence (the Hill sphere radius Eq. 12.5). The total mass of planetesimals in this *feeding zone* is therefore

$$m_{\mathrm{fz}} = 2\pi d\,2r_{\mathrm{H}}\,\Sigma_{\mathrm{p}} \tag{12.13}$$

The maximum attainable mass is reached by accreting all of the planetesimals in the feeding zone and is referred to as the *isolation mass*. It is therefore given by

$$m_{\mathrm{iso}} = 2\pi d\,2r_{\mathrm{H}}(m_{\mathrm{iso}})\,\Sigma_{\mathrm{p}} = 4\pi d^2\left(\frac{m_{\mathrm{iso}}}{3M_\star}\right)^{1/3}\Sigma_{\mathrm{p}}. \tag{12.14}$$

This leads to

$$m_{\mathrm{iso}} = \left(\frac{4\pi}{3^{1/3}}\right)^{2/3}\Sigma_{\mathrm{p}}^{3/2}\,d^3\,M_\star^{-1/2}. \tag{12.15}$$

For disk models that have surface density structures as expected, such as Eq. 9.31, $m_{\mathrm{iso}}$ is an increasing function of the orbital radius $d$, so we expect more massive solid bodies to form at larger orbital radii. Evaluating this expression in the terrestrial planet region, taking $d = 1\,\mathrm{AU}$, $\Sigma_{\mathrm{p}} = 100\,\mathrm{gr\,cm}^{-2}$, $M = M_\odot$, we obtain $r_{\mathrm{H}} = 0.01\,\mathrm{AU}$ and $m_{\mathrm{iso}} = 0.05\,M_\oplus$. Therefore, isolation is likely to occur late in the formation of the terrestrial planets. Indeed, the assembly of terrestrial planets probably requires the agglomeration of quite a number of isolation-mass objects in a single planet. Models suggest that terrestrial planet formation via repeated impacts of isolation-mass objects takes $\sim\!100\,\mathrm{Myr}$, which is broadly consistent with Solar System observations.

An estimate of the isolation mass at the location of the giant planets, at $d = 5 - 30\,\mathrm{AU}$, using $\Sigma = 1 - 10\,\mathrm{gr\,cm}^{-2}$, results in a few $M_\oplus$ for the isolation mass $m_{\mathrm{iso}}$. This is somewhat below the current best determinations for the mass of the Jovian core, however, more realistic

estimates tend to produce isolation masses that are somewhat larger – typically $\gtrsim 10\,M_\oplus$. This is comparable to (or greater than) the core masses of the Solar System giant planets. Full isolation may or may not be relevant to the formation of Jupiter and the other giant planets, depending on the adopted disk model.

## Oligarchic growth

Runaway growth ends when the largest proto-planets dominate the dynamics of the planetesimal disk. When these runaway bodies, or planetary embryos, become sufficiently massive, it is their gravitational scattering (often called viscous stirring) which dominates the random velocity evolution of the background planetesimals, rather than the weak gravitational encounters among the planetesimals (called viscous self-stirring). The boost in viscous stirring that the planetary embryos provide increases the random energies of the planetesimals and the equilibrium values of planetesimal eccentricity and inclination, hence their velocities relative to the embryos. Since the accretion cross-section of a planetary embryo is smaller among planetesimals with higher velocity (see Eq. 12.4) proto-planet growth now switches to a slower, self-limiting mode. The subsequent accretion mode has been given the name *oligarchic growth*. The large bodies still outgrow planetesimals, but each grow at a similar rate and with roughly even spacing, eventually forming proto-planets. So, across the disk many *oligarchs* grow at similar rates by consuming planetesimals within their own largely independent feeding zones. The total number of planetary embryos that form in the inner disk is uncertain, and may range from $\sim 30 - 50$ if these are relatively massive (say, $10^{-1}\,M_\oplus$) to perhaps $500 - 1000$ if embryos average only a lunar mass (Raymond et al. 2006). The timescale for embryo formation is thought to increase with orbital radius $d$, and decrease with the local surface density. Thus, embryos form quickly at small orbital radii and slower farther from the central star.

At the *snow line* or frost line, where the temperature drops below the condensation point of water (at about $150-170$ K; e.g. Stevenson & Lunine 1988), the local density of solid particles increases rather abruptly (see Fig. 12.5). This jump in density of a factor of $\sim 3 - 4$ changes the time- and mass-scales of planets that form beyond this distance: the isolation mass $m_{\mathrm{iso}}$ of planetesimals increases by a factor of $\sim 5 - 8$, and these cores also form faster. These massive icy planetary embryos are thought to be the building blocks of giant planet cores, following the *core-accretion scenario* for giant planet formation (Pollack et al. 1996). Thus, planetary embryo formation is different interior and exterior to the snow line – the longest embryo formation timescales may lie in the outer regions of the inner disk (Kokubo & Ida 2002).

## Final assembly

Gravitational scattering *among* planetesimal bodies of the same mass results in a steady increase of the random component of the planetesimal velocities, and a corresponding thicken-
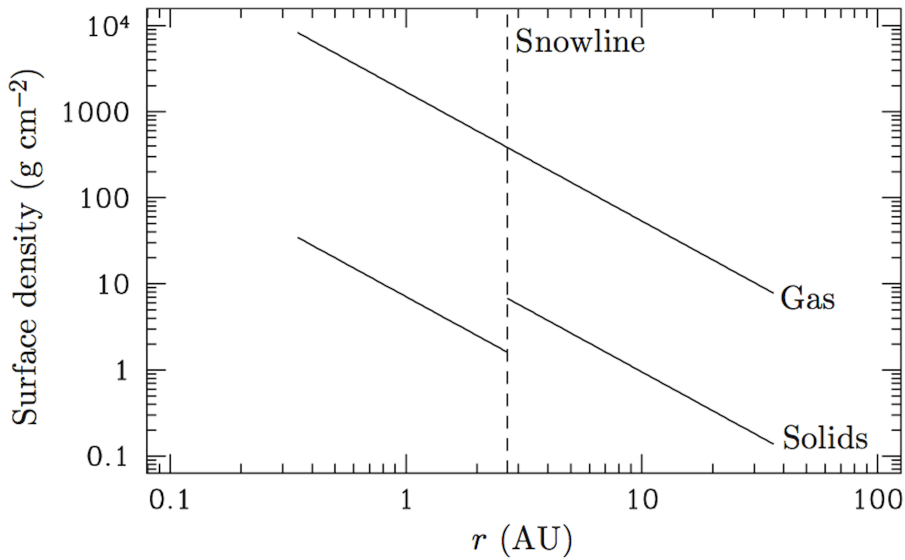
Figure 12.5: *The surface density in gas (upper line) and solids (lower broken line) as a function of radius in the minimum mass Solar Nebula of Hayashi (1981). The dashed vertical line denotes the location of the snow line at $d = 2.7\,AU$, where $T = 170\,K$. Later studies place the snow line somewhat further out, at $d = 3 - 3.2\,AU$ and temperatures of $143 - 150\,K$ (Podolak & Zucker, 2010). From: Hayashi C, 1981, Progress of Theoretical Physics 70, 35.*

ing of the planetesimal disk – as pointed out above[1]. Once the oligarchs form, a new effect arises which derives from the fact that an encounter between a low mass body of mass $m$ and a larger one of mass $M$ gives a greater impulse to the lower mass object. The system tries to attain a state of energy equipartition in which

$$\frac{1}{2}m\,\sigma_m^2 = \frac{1}{2}M\sigma_M^2, \tag{12.16}$$

where $\sigma_m$ and $\sigma_M$ are the random velocities of the low and high mass bodies. This process is known as *dynamical friction*, and it leads to a mass-dependence of the mean eccentricity and inclination of planetesimals and growing proto-planets. Notice that the relatively small velocities of the massive objects in itself is beneficial for gravitational focussing, hence in this sense dynamical friction tends to aid in the (runaway) growth of the larger bodies.

The final assembly of terrestrial planets gets underway once the oligarchs have depleted the planetesimal disk to the point that dynamical friction can no longer maintain low velocities and low eccentricities and inclinations among the oligarchs. The assumption that each oligarch grows in isolation breaks down, and the largest bodies start to interact strongly, collide, and scatter smaller bodies across a significant radial extent of the disk (Armitage 2010).

---

[1]The effect that counters this viscous stirring is that of aerodynamic drag. Although much weaker for planetesimals than for meter-sized bodies, it damps the effect of viscous stirring on both eccentricity and inclination. Actually, the balance between stirring and aerodynamic drag sets the distributions of $e$ and $i$.
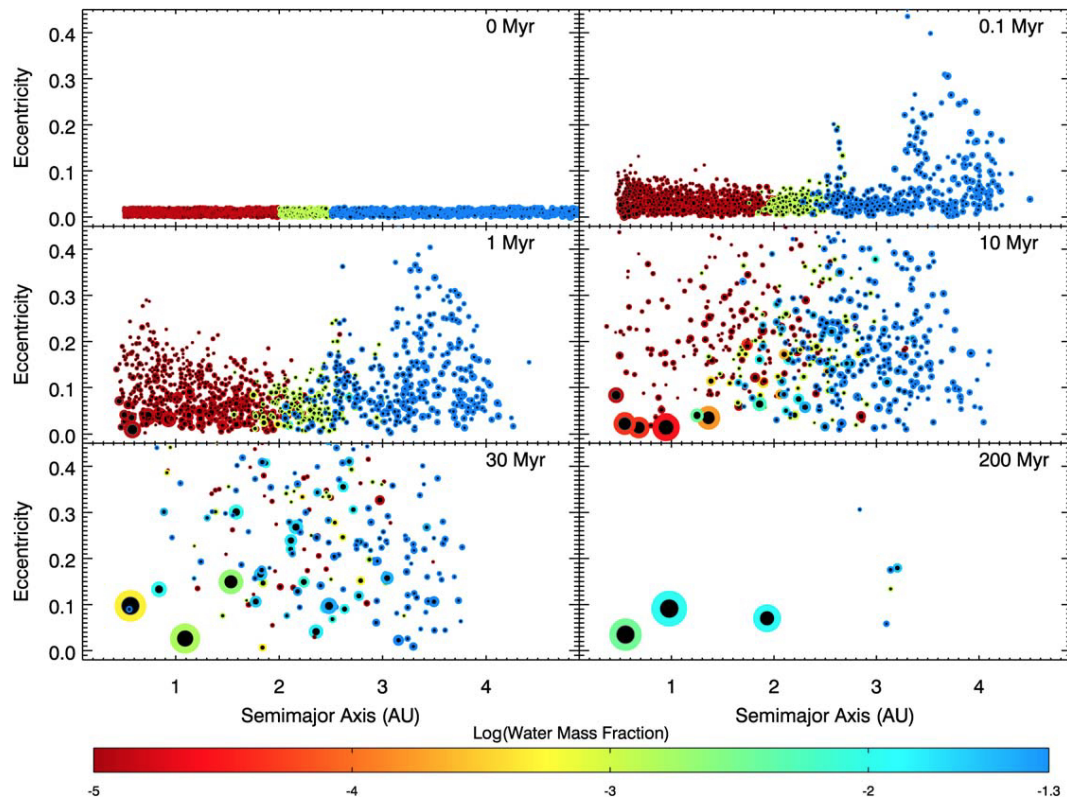
Figure 12.6: *Six snapshots in time from an N-body simulation of the final assembly of terrestrial planets by Raymond et al. (2006). The initial conditions of this simulation are representative of the late stages of oligarchic growth, with a grand total of 1185 planetesimals and somewhat lager bodies initially distributed within the range of terrestrial planet formations (up to 5 AU). The size of each body corresponds to its relative physical size (i.e., its mass $M^{1/3}$), but is not to scale on the $x$-axis. The color of each particle represents its water content, and the dark inner circle represents the relative size of its iron core. There is a Jupiter-mass planet at 5.5 AU on a circular orbit (not shown). See main text for further details. From: Raymond et al. 2006, Icarus 183, 265.*

## 12.1.1 Raymond et al. 2006 simulation of the final phase of terrestrial planet formation

Numerical N-body simulations (e.g. Raymond et al. 2006; O'Brien et al. 2006; Mandell et al. 2007) probe this final phase in the formation of terrestrial planets. As an example, we discuss simulation 0 from Raymond et al. 2006 – shown in Fig. 12.6). It is important to realize that the proces of dynamical interactions is chaotic, and hence identical initial conditions can give rise to a range of outcomes that must be compared to the observed properties of terrestrial planetary architectures in a statistical way. In this simulation the aim is to reproduce the Solar System's configuration, however, only one simulation for the set of initial conditions has been performed.

The adopted surface density profile is $\Sigma_{\rm p} = 10\,(r/1\,{\rm AU})^{-3/2}\,{\rm gr\,cm}^{-2}$. The total mass of solid bodies is $9.9\,M_\oplus$, extending to 5 AU. All (at first 1885) planetesimals are given small initial eccentricities ($e \leq 0.02$) and inclinations ($i \leq 1°$). Inside 2 AU bodies are initially dry; between 2 and 2.5 AU they contain 0.1% water by mass, and outside of 2.5 AU they have an initial water content of 5% by mass. The simulation starts in the late stages of oligarchic growth, when planetary embryos were not yet fully formed. The separation between embryos is randomly chosen to lie between 0.3 and 0.6 *mutual Hill radii*

$$r_{\rm H,mutual} = \left(\frac{m_1 + m_2}{3\,M_\odot}\right)^{1/3} \left(\frac{d_1 + d_2}{2}\right),\qquad (12.17)$$

where $d_1$ and $d_2$ are the two embryo's orbital distances, and $m_1$ and $m_2$ their masses. The stellar mass is $M_\odot$. A Jupiter-mass giant planet on a circular orbit is present at 5.5 AU. It is thus assumed that Jupiter is constrained to have formed in the few Myr lifetime of the gaseous component of the solar nebula. In the initial state only sub embryo-sized bodies are present, therefore this simulation implies a very fast timescale for Jupiter's formation. The smallest bodies are $10^{-3}\,M_\oplus$, with some even less massive in the inner edge of the simulation. Collisions were treated as inelastic mergers conserving water and iron. The evolution of the system was followed for 200 Myr, i.e., it is by far the slowest stage of planet formation.

0.1 Myr snapshot. Planetesimals and larger embryos quickly (already in the 0.1 Myr snapshot) become excited by their own mutual gravitation as well as that of the Jupiter-mass planet at 5.5 AU (which henceforth we will call Jupiter). Objects in several distinct *mean motion resonances* have been excited to relatively extreme eccentricities: the 3:1 resonance at 2.64 AU, 2:1 resonance at 3.46 AU, and the 5:3 resonance at 3.91 AU. Such resonances occur when two orbiting bodies exert a regular, periodic gravitational influence on each other, usually because their orbital periods are related by a ratio of two small integers. Think of the concept of a parent pushing a child on a swing, where the swing has a natural frequency (or period), and the parent doing the 'pushing' will act in periodic repetition to have a sustaining effect on the motion. Likewise, orbital resonances greatly enhance the mutual gravitational influence of the bodies, i.e. their ability to alter or constrain each other's orbits. In most cases, this results in unstable interactions, in which the bodies exchange momentum and shift orbits until the resonance no longer exists. An example of this is also visible in the 0.1 Myr snapshot: all material exterior to the 3:2 resonance at 4.2 AU is quickly removed from the system via interactions with and ejections by the giant planet at 5.5 AU. Under some circumstances, however, a resonant system can be stable and self-correcting, so that the bodies remain in resonance (in this case the 3:1, 2:1, and 5:3 resonance).

1 Myr snapshot. Once eccentricities are large enough for the orbits of particles to cross, bodies begin to grow via accretionary collisions. At 1 Myr this has started to happen in the inner disk. The large bodies tend to have small eccentricities and inclinations, due to the dissipative effects of dynamical friction (see Eq. 12.16).

10 Myr snapshot. Accretion proceeds fastest in the inner disk and moves outward, as can be seen by the presence of several large bodies inside 2 AU after 10 Myr. In the outer disk,
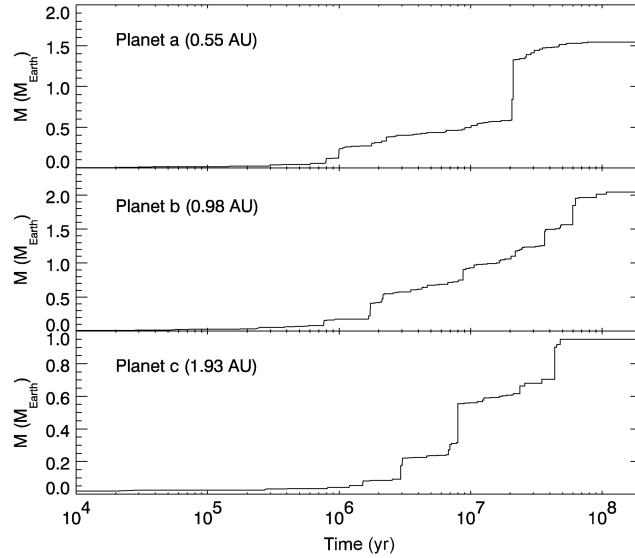
Figure 12.7: *Mass evolution for the three terrestrial planets from simulation 0 in Raymond et al. (2006).*

collisions are much less frequent because of the slower dynamical timescale and decreasing density of material due to Jupiter clearing the region via ejections. However, the mean eccentricity of bodies in the outer diks is quite high, enabling those which are not ejected by Jupiter to interact with proto-planets closer to the sun. Multiple encounters may cause the orbits of asteroidal proto-planets to move inward in time (i.e., their semi-major axes can decrease).

30 Myr snapshot. After 30 Myr the inner disk is composed primarily of a few large bodies, which have begun to accrete water-rich material from the outer disk. Larger bodies have started to form out to $2.5 - 3$AU, but their eccentricities are large.

200 Myr. In time, several of the outer large bodies are incorporated into the three final terrestrial planets, defined as those objects that reach a mass $> 0.2 \, M_\oplus$. Their properties are listed in Table 12.2.

| Planet | $a$ AU | $e$ | $i$ | $M$ $M_\oplus$ | $N_\mathrm{p}$ | #coll |
|--------|------|------|-----|------|------|------|
| $a$ | 0.55 | 0.05 | 2.8 | 1.54 | 500 | 87 |
| $b$ | 0.98 | 0.04 | 2.4 | 2.04 | 457 | 98 |
| $c$ | 1.93 | 0.06 | 4.6 | 0.95 | 174 | 174 |

Table 12.2: *Properties of the outcome of simulation 0 of Raymond et al. (2006).*
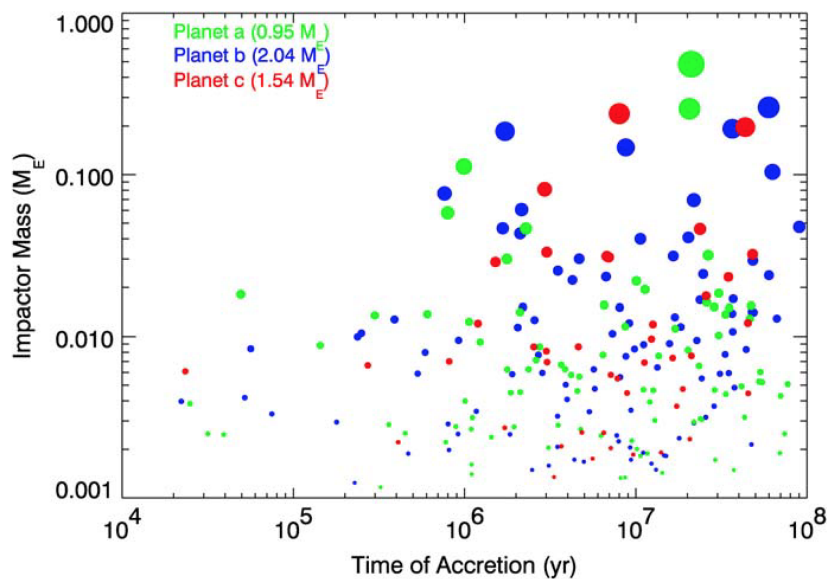
Figure 12.8: *Impactor mass versus time for all bodies accreted by the surviving planets in simulation 0 in Raymond et al. (2006). The size of each body is proportional to its mass $M^{1/3}$, and represents its relative physical size.*

Let us focus a bit on planet $b$, as out of the three this one is in the habitable zone. It accumulated its final mass of 2.04 $M_{\oplus}$ from 487 protoplanets via 98 accretionary collisions. The feeding zone of this planet – the zone of starting points of all bodies which were incorporated into that planet – extends past 4 AU, though a mere 25% (17%) of the bodies originate from beyond 2 (2.5) AU. These, i.e. bodies from $> 2$ AU, are almost exclusively accreted relatively late, after 10 Myr. Figure 12.7 shows the mass of each of the three final planets as a function of time. Each planet grows quickly in the first few million years, reaching half of its final mass within $10 - 20$ Myr, after which the growth rate flattens off. The later stages of growth are characterized by a step-wise pattern representing a smaller number of larger-scale collisions with other oligarchs, which have also cleared out all bodies in their feeding zones. This is further clarified in Fig. 12.8, which shows the mass of bodies accreted by each of the surviving planets as a function of time. Bodies of a range of sizes are accreted by each planet throughout the simulation, although large impacts tend to happen later in the simulation. Such large, late impacts are reminiscent of the Moon-forming impact (e.g. Canup & Asphaug, 2001). The final body accreted by planet $b$ occurs at slightly after 100 Myr.

## 12.2  Giant planet formation

Two qualitatively different theories have been proposed to account for the formation of massive planets: the *core accretion* mechanism (Chapters 10 and 12) and the *gravitational instability* mechanism (see Sect. 9.5). In the *core accretion* mechanism a rocky and icy core forms as described above, followed by the acquisition of a massive envelope of gas. Given the typical lifetime of gaseous planetary disks of $3 - 10\,\mathrm{Myr}$ (see Fig. 9.13), the timescale for giant planet formation in this model must be relatively short – it hinges on how quickly the core can be assembled and on how fast the gas in the envelope can cool and accrete on to the core. In the competing disk instability theory, giant planets form promptly via the gravitational fragmentation of an unstable gas-rich proto-planetary disk, as described in Sect. 9.5. Beware that these two theories need not be mutually exclusive. Both mechanisms may operate, potentially even around the same star.

We focus here on probably the main route toward giant planet formation, i.e. core accretion. The key to understanding core accretion lies in the behaviour of the gaseous envelope as the core mass grows. A core of several Earth masses moving through the gas of the proto-planetary disk has strong enough gravity to retain a dense atmosphere. This atmosphere is initially in hydrostatic equilibrium, meaning that the inward force of gravity on the envelope (coming primarily from the gravity of the solid core) is balanced by an outward pressure gradient that is set by a temperature and density gradient (see Eqs. 3.2, 5.1 and 5.2). Perhaps somewhat surprisingly, there is a maximum core mass above which it is not possible to have an envelope in hydrostatic equilibrium. For simplicity, assume that the gaseous envelope is isothermal, i.e. the pressure gradient is solely the result of a density gradient. Once $M_{\mathrm{env}} \sim M_{\mathrm{core}}$ a situation develops in which a further increase of the envelope mass, though providing extra density for pressure gradient support, is no longer able to keep up the balance with gravity as the very same material is increasing the gravitation force by too much. For higher core masses no hydrostatic solution exists, since any possible boost to the pressure by adding yet more gas to the envelope fails to compensate for the additional mass. The critical core mass for this to happen is typically of the order of $10\,M_{\oplus}$.

One distinguishes four main phases in the formation of giant planets (see Fig. 12.9). They are:

- Core formation. Formation of a solid proto-planet (henceforth 'core'), in the way as described in the discussion of terrestrial planet formation, until it reaches a mass large enough to retain a significant gaseous atmosphere or envelope. The rate of growth of this core is determined by the initial surface density of rocky and icy bodies and by the extent of gravitational focussing.

- Hydrostatic growth. This phase is characterized by the slow accretion of gas from the proto-planetary disk. Initially the envelope surrounding the solid core is in hydrostatic equilibrium. Energy liberated by planetesimals impacting the core, together with gravitational potential energy released as the envelope itself contracts, must be transported
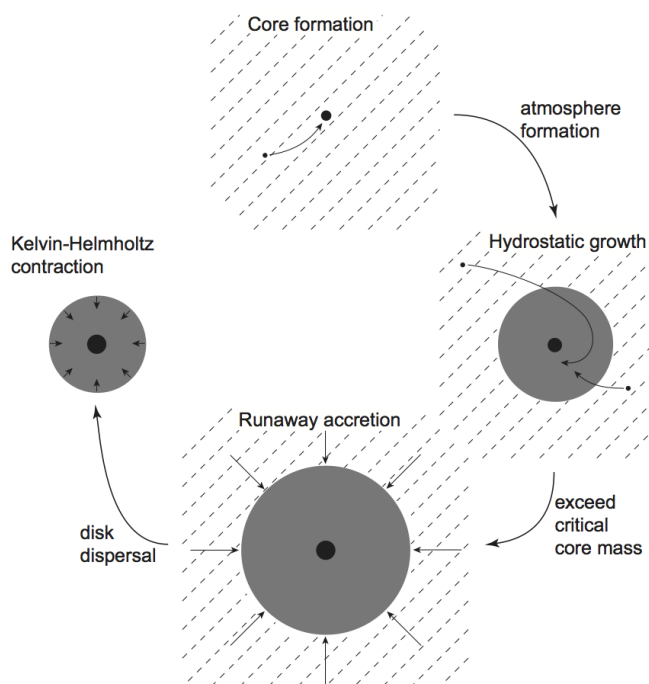
Figure 12.9: *The main stages in the formation of giant planets via core accretion. Once the accretion has terminated the planet enters the evolution phase of cooling and quasi-hydrostatic contraction. From: Philip J. Armitage, Astrophysics of Planet Formation (2010).*

through the envelope by either radiative diffusion or convection and is emitted at infrared wavelengths from the surface. The core and envelope grow in mass.

- Runaway growth. Once the critical mass is exceeded a runaway phase of gas accretion ensues in which the envelope increases in mass much more rapidly than the core. The rate of growth is only limited by the supply of gas in the planet's feeding zone in the proto-planetary disk. For massive planets the bulk of the planetary envelope is accreted during this phase, which lasts on the order of $10^5$ yr.

- Termination of accretion. Eventually the supply of gas is exhausted, either as a consequence of the dissipation of the entire proto-planetary disk (see Sect. 9.6) or as a consequence of the gas giant opening up a local gap in the disk. Accretion of material subsides and the planet 'detaches' from the nebula.

Once accretion of material has seized, the planet commences a long phase of cooling and quasi-hydrostatic contraction often referred to as the evolution phase.

### The start of atmosphere accretion

A solid body can start to retain gas when its surface escape velocity $v_{\mathrm{esc}} = \sqrt{2GM_{\mathrm{c}}/R_{\mathrm{c}}}$ exceeds the local sound speed $a$ of the gas in which it is embedded. This is known as *Bondi accretion*. Assuming a material density $\rho_{\mathrm{m}}$, such that $M_{\mathrm{c}} = (4\pi/3)\,R_{\mathrm{c}}^3\,\rho_{\mathrm{m}}$, and using that the sound speed $a = (h/r)\,v_{\mathrm{K}}$ (see Eq. 9.26), one obtains for the condition $v_{\mathrm{esc}} > a$

$$M_{\mathrm{c}} > \left(\frac{3}{32\pi}\right)^{1/2} \left(\frac{h}{r}\right)^3 \frac{M_{\star}^{3/2}}{\rho_{\mathrm{m}}^{1/2}\,r^{3/2}}, \tag{12.18}$$

where $M_{\star}$ is the mass of the central star. At a 1 AU distance from a solar mass star, assuming a material density of $3\,\mathrm{gr\,cm}^{-3}$ and $h/r = 0.05$, one finds that $M_{\mathrm{c}}$ should be at least $3 \times 10^{-3}\,M_{\oplus}$. At 5 AU, where $\rho_{\mathrm{m}}$ is more likely to be $1\,\mathrm{gr\,cm}^{-3}$ one obtains $M_{\mathrm{c}} > 5 \times 10^{-4}\,M_{\oplus}$.

One may equivalently formulate the criterium for the start of atmosphere formation in terms of the *Bondi radius*

$$R_{\mathrm{Bondi}} = \frac{2GM_{\mathrm{c}}}{a^2}. \tag{12.19}$$

The planet starts to hold on to an atmosphere once the Bondi radius becomes larger than the physical radius. For the same examples as worked out above we find that at 1 AU, where $a = 1.5\,\mathrm{km\,s}^{-1}$, $R_{\mathrm{Bondi}} = 1150\,\mathrm{km}$. At 5 AU, where $a = 0.7\,\mathrm{km\,s}^{-1}$, it is $900\,\mathrm{km}$.

So, already a relatively small body (an embryo) may be able to assemble a gaseous atmosphere. However, realize that an object that marginally satisfies the Bondi condition would lose its atmosphere on approximately the dynamical time scale in the absence of the gas disk, so this is *not* the same condition as for an isolated planet to be able to retain an atmosphere. A planet that is not embedded in a gas must have a surface escape speed that is much higher than the local sound speed to be able to retain the molecules that have speeds far out in the Maxwellian tail of the particle velocity distribution – a process known as Jeans escape.

### Retaining a substantial gaseous envelope

What is the minimum core mass required to maintain an envelope with a mass that is a non-negligible fraction of the mass of the core? To estimate this mass we assume that the envelope is isothermal, such that $p = \rho\,a^2$, where $a$ is the sound speed. Using this expression for the pressure, the equation of hydrostatic equilibrium for the envelope can be written as

$$\frac{\mathrm{d}p}{\mathrm{d}r} = a^2\frac{\mathrm{d}\rho}{\mathrm{d}r} = -\frac{GM_{\mathrm{p}}}{r^2}\,\rho, \tag{12.20}$$

such that

$$\frac{1}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}r} = -\frac{GM_{\mathrm{p}}}{a^2}\frac{1}{r^2} = -\frac{R_{\mathrm{Bondi}}}{2}\frac{1}{r^2}, \tag{12.21}$$

hence

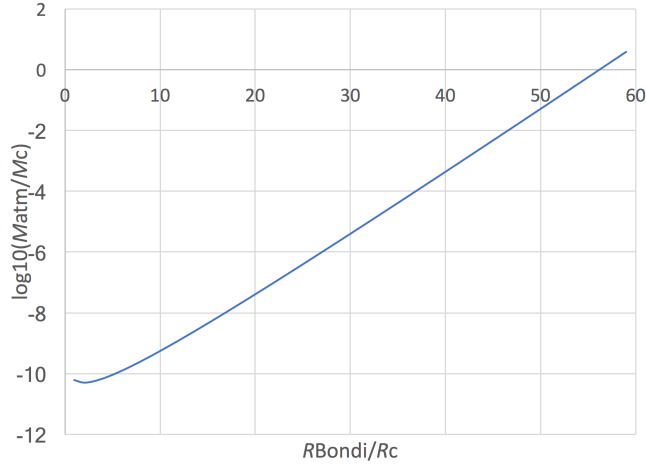$$\ln\rho = \frac{R_{\mathrm{Bondi}}}{2}\frac{1}{r} + \text{constant}. \tag{12.22}$$

Figure 12.10: *Atmosphere-mass to core-mass ratio as a function of $R_{\mathrm{Bondi}}/R_{\mathrm{c}}$ for a density ratio $\rho_{\mathrm{o}}/\rho_{\mathrm{m}} = 1 \times 10^{-11}$.*

To evaluate the constant on the right-hand-side we require that at the Bondi radius the envelope density $\rho$ matches the density $\rho_{\mathrm{o}}$ in the unperturbed disk, i.e. we assume pressure equilibrium at $R_{\mathrm{Bondi}}$. This yields

$$\rho(r) = \rho_{\mathrm{o}} \exp\left\{\frac{1}{2}\left[\frac{R_{\mathrm{Bondi}}}{r} - 1\right]\right\}. \tag{12.23}$$

The density at the disk mid-plane $\rho_{\mathrm{o}}$ may be related to the surface density and vertical scale-height through Eq. (9.27).

The pressure scale height corresponding to Eq. (12.23) is $H_{\mathrm{p}} = 2r^2/R_{\mathrm{Bondi}}$. Most of the mass of an envelope with the above density profile lies in a relatively thin shell close to the surface of the solid core. Let us assume this shell to have a thickness of one pressure scale height, such that $H_{\mathrm{p}}/r = 2r/R_{\mathrm{Bondi}} << 1$. In that case the mass of the atmosphere can be approximated as

$$
\begin{aligned}
M_{\mathrm{atm}} &\simeq 4\pi R_{\mathrm{c}}^2 H(R_{\mathrm{c}}) \rho(R_{\mathrm{c}}) \\
&= 8\pi R_{\mathrm{Bondi}}^3 \rho_{\mathrm{o}} \left(\frac{R_{\mathrm{c}}}{R_{\mathrm{Bondi}}}\right)^4 \exp\left\{\frac{1}{2}\left[\frac{R_{\mathrm{Bondi}}}{R_{\mathrm{c}}} - 1\right]\right\},
\end{aligned}
\tag{12.24}
$$

where $R_{\mathrm{c}}$ is the core radius. For a material density $\rho_{\mathrm{m}}$ of the core, this may be rewritten as

$$\frac{M_{\mathrm{atm}}}{M_{\mathrm{c}}} = 6 \left(\frac{\rho_{\mathrm{o}}}{\rho_{\mathrm{m}}}\right) \left(\frac{R_{\mathrm{c}}}{R_{\mathrm{Bondi}}}\right) \exp\left\{\frac{1}{2}\left[\frac{R_{\mathrm{Bondi}}}{R_{\mathrm{c}}} - 1\right]\right\}. \tag{12.25}$$

To put real numbers to this exercise, let us consider an icy body ($\rho_{\mathrm{m}} = 1\,\mathrm{gr\,cm}^{-3}$) growing at 5 AU, where $\rho_{\mathrm{o}} = 10^{-11}\,\mathrm{gr\,cm}^{-3}$ and $a = 7 \times 10^4\,\mathrm{m\,s}^{-1}$. Figure 12.10 shows the ratio $M_{\mathrm{atm}}/M_{\mathrm{c}}$ as a function of $R_{\mathrm{Bondi}}/R_{\mathrm{c}}$ for this ratio $\rho_{\mathrm{o}}/\rho_{\mathrm{m}}$. We may express $M_{\mathrm{c}}$ in terms of

the ratio $R_c/R_{Bondi}$. This yields

$$M_c = \left(\frac{3}{32\pi\rho_m}\right)^{1/2} \left(\frac{a^2}{G}\right)^{3/2} \left(\frac{R_{Bondi}}{R_c}\right)^{3/2}. \tag{12.26}$$

If we assume that $M_{atm} > 0.1\,M_c$ marks the threshold above which the envelope atmosphere could be said to be significant, i.e. at $R_{Bondi}/R_c = 56.1$, we predict that this will occur for core masses $M_c \gtrsim 0.24\,M_\oplus$. At 1 AU a similar estimate would yield a core mass $M_c \gtrsim M_\oplus$ before we would expect it to be enshrouded in a dense massive envelope.

Recall the discussion on the isolation mass Eq. (12.14) and note that at 1 AU the core mass required to hold a sizeable atmosphere is *larger than* the isolation mass. In these regions it is therefore unlikely that protoplanets will grow fast enough to capture envelopes prior to disk dispersal, and the ultimate outcome of planet formation will instead be terrestrial planets. At 5 AU the situation is reversed, a core mass *less than* the isolation mass is sufficient to retain a significant envelope. Giant planet formation is much more probable in this case.

Ultimately, the growth of the atmosphere stops. Potential reasons for this may include:

- The atmosphere does not extend all the way to $R_{Bondi}$. Certainly, at $r = r_H$ (stellar) tidal forces become important and it is inconceivable that planets can bind material beyond the Hill radius. This will likely limit the atmosphere size for large planets, once $R_{Bondi}$ has exceeded $r_H$. Nevertheless, as $r_H/R_c \gg 1$ proto-planetary atmospheres that fill the Hill sphere are still enormous.

- The planet accretes the isolation mass and opens up a local *gap* in the disk.

- The entire gaseous proto-planetary disk is dissipated after $3 - 10$ Myr (see Fig. 9.13).

Once accretion tails off the gas giant commences a long phase of cooling and quasi-hydrostatic contraction.

### 12.2.1  Mordasini et al. 2012 simulation of the formation and evolution of Jupiter

Mordasini et al. (2012) present a simulation of the *in situ* formation of a Jovian mass planet at 5.2 AU using the same initial planetesimal surface density, disk pressure, and temperature as Pollack et al. (1996), a case that has been studied frequently in the literature. Though complex enough, the simulation is strongly simplified in several respects: *(i)* The formation of the planet and the evolution of the disk are not treated self-consistently. If one would, the disk-limited runaway gas accretion rate and the moment of the onset of limited gas accretion, consequently the final mass of the planet, would naturally result from the evolution of the gaseous disk. Here, the runaway gas accretion rate is pre-specified, limited to a maximum of $10^{-2}\,M_\oplus$/yr, and artificially ramped down on a short timescale once the mass of the planet

reaches 1 $M_{\mathrm{Jup}}$. *(ii)* Migration of the Jovian proto-planet is completely switched-off. If considered, it would be a central component of the formation model as it can lead to a much more complex accretion history (e.g. Mordasini et al. 2009, A&A 501, 1139).

The model further assumes that all energy liberated in the accretion shock on the planetary surface is radiated away from the planet, hence does not contribute to the luminosity inside the planet's structure. This situation, where the gravitational energy released by newly accreted material is lost as $L_{\mathrm{acc}}$, is referred to as the *cold start scenario*. The opposite limiting case is to assume that no energy at all is radiated away in the shock, referred to as the *hot start scenario*. The latter is regarded a physically unlikely scenario.

Figure 12.12 shows the most important planetary properties as function of time during the formation process and at the very beginning of the evolutionary phase once the final mass is reached. The main phases in the formation process and the beginning of the evolution phase are identified in the left upper panel using thin grey (and hard to spot) vertical lines and the labels I, II, III, D, and E. The initial core formation phase (phase I) is characterised by the rapid build-up of the core until the isolation mass is reached. Then a hydrostatic phase follows characterised by a slow increase of the envelope mass, which allows further core growth, and which ends when runaway growth sets in.

The crossover point (between phase II and III) when the core and envelope masses are the same is reached at 0.82 Myr. The duration of the hydrostatic phase critically depends on the dust opacities in the planet's gaseous envelope, which are affected by grain growth and settling. In general, the opacities are substantially lower than interstellar values and (compared to interstellar values; see Fig. 6.4) result in more rapid heat loss from and more rapid contraction of the proto-planetary envelope. This yields a substantially shorter phase II formation phase. The mass of the core at 0.82 Myr is $M_{\mathrm{core}} = 16.3\,M_{\oplus}$. Shortly after reaching the crossover point, the gas accretion rate increases strongly quickly hitting the maximal allowed value of $10^{-2}\,M_{\oplus}$/yr at 0.923 Myr. The total mass of the planet is then 86.7 $M_{\oplus}$, and $\tau_{\mathrm{KH}}$ is as short as $\sim 1640$ yr. As discussed in Sect. 6.4, this implies that in order to produce its internal luminosity $L_{\mathrm{int}}$ the planet will shrink quickly. The rapid contraction, which is assumed to be a hydrostatic contration (but see Wuchterl 1991) is referred to as the collapse of the envelope. Shortly afterwards, at about 0.961 Myr, the final mass of the planet is almost reached (99.5% of the final mass to be precise). This marks the beginning of the evolutionary phase, labeled with E.

The model considers the accretion of planetesimals. This accretion also increases strongly during the runaway phase because of the fast expansion of the planet's feeding zone due to the increase of the Hill radius. It reaches a maximum of about $7 \times 10^{-4}\,M_{\oplus}$/yr and then starts to decrease *before* it is ramped down artificially. The reason for the early turn-down is that ejection of planetesimals becomes increasingly important as the mass of the planet grows and because the capture radius of the planet shrinks (see dotted-green line in bottom left panel of Fig. 12.12). The final total mass of the planet is 316.6 $M_{\oplus}$. Assuming that all accreted planetesimals can reach the core, the final core mass is 32.9 $M_{\oplus}$. It is more likely however that once the envelope mass has grown above a certain mass (about $1 - 2\,M_{\oplus}$ for 100 km
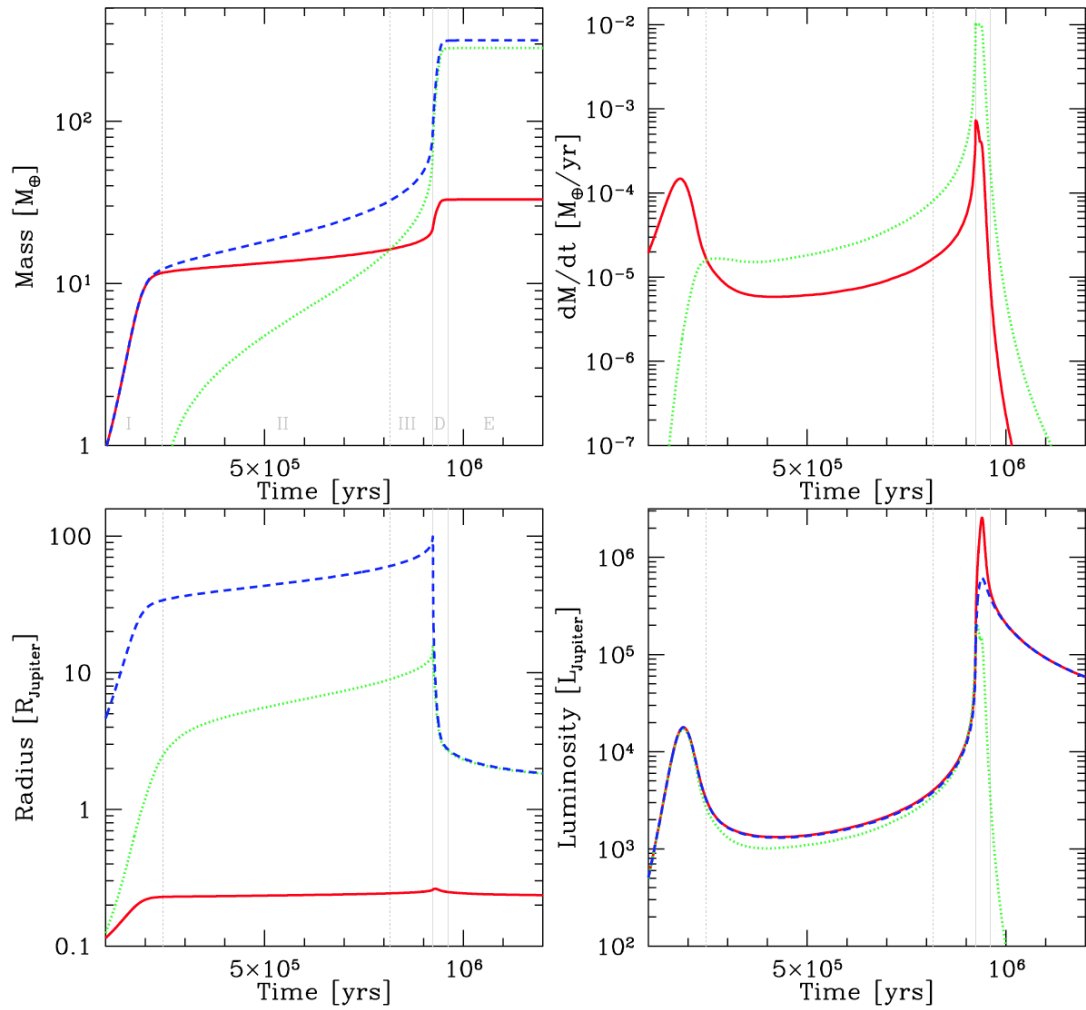
Figure 12.11: *Simulation of the in situ formation of Jupiter at 5.2 AU, so without taking into account possible migration during formation. Gray (difficult to see) vertical lines show the major phases, labeled in the top left panel: I (core formation), II (hydrostatic growth), III (runaway growth), D (termination of accretion), and E (evolution). Top left: evolution of the core mass (red solid line), the envelope mass (green dotted line), and the total mass (blue dashed line). Top right: accretion rate of solids (red solid line) and gas (green dotted line). The limiting gas accretion rate is fixed to $10^{-2} M_\oplus$/yr. Bottom left: evolution of the core radius (red solid line), the total radius (blue dashed line), and the capture radius for planetesimals (green dotted line). The capture radius is found by integrating the orbits of planetesimals inside the envelope (see Alibert et al. 2005, A&A 434, 343). Bottom right: luminosity of the planet in present day intrinsic luminosities of Jupiter ($L_{\rm Jup} = 8.7 \times 10^{-10} L_\odot$). The red solid line is the total luminosity $L$, the blue dashed line the internal luminosity $L_{\rm int}$ and the green dotted line is the core luminosity $L_{\rm core}$. It follows that the accretion luminosity is $L_{\rm acc} = L - L_{\rm int}$. From: Mordasini et al. 2012, A&A 547, A111.*

planetesimals; Mordasini et al. 2005), planetesimals cannot penetrate to the core of the planet but get destroyed in the envelope. So, it appears more appropriate to associate the 32.9 $M_\oplus$ value with the total mass of heavy elements contained in the planet rather than the core mass, though over time (part of) this material may sink to the core-envelope interface. Still, when the envelope mass reaches 2 $M_\oplus$, the core mass is about 12 $M_\oplus$. This value may be considered to represent the minimum core mass.

The bottom left panel of Fig. 12.12 shows the evolution of the total radius, the core radius, and the capture radius. The latter is found by integrating the orbits of planetesimals inside the orbit. In phase II the total radius is approximately one-third of the Hill sphere radius. The capture radius during this phase is much larger than the core radius, which is very important for the core growth. The envelope is characterised by very large scale heights of thousands to several ten thousands of kilometers, where not all planetesimals are actually captures. This is why the capture radius is considerably smaller than the total radius. After the collapse the envelope structure becomes much more compact and the planet develops a well-defined surface (i.e. a much smaller scale height near the surface) at which planetesimals are captures. Even though the core grows in mass considerably in phase II, the core radius remains about constant. The reason is that the external pressure on the core by the accreted gas increases so much that the resulting strong increase of the core density over-compensates the mass increase.

The bottom right panel shows the core luminosity $L_{\rm core}$, the total luminosity $L$, and the luminosity within the planetary structure $L_{\rm int}$. The difference $L - L_{\rm int}$ is thus the shock/accretion luminosity $L_{\rm acc}$. The first peak in the curve is due to the rapid accretion of the core and the second to the combined effect of runaway gas accretion and envelope collapse. The second luminosity maximum occurs at 0.94 Myr, with $L = 2.6 \times 10^6 \, M_{\rm Jup} = 0.0022 \, L_\odot$. This is about the luminosity of an M5 V star. At this point accretion contributes most to the total luminosity and may be larger than $L_{\rm int}$ by up to a factor of $\sim 3$.

Once accretion has stopped, the planet evolves at constant mass. The evolution phase starts as the planet shrinks and becomes dimmer over time. After 4.6 Gyr, it has a radius of 0.99 $R_{\rm Jup}$ and a luminosity of 1.13 $L_{\rm Jup}$. Indeed, we have essentially been discussing the formation of our Jupiter.

## 12.3 Mass-Radius relation of planets

As these lecture notes are concerned only with the formation process of planets, this is the opportune moment to wrap up. Still, we add one more topic: the mass-radius relation, particularly that of terrestrial planets – those that are composed of rock or some other (combination of) solid material and that are surrounded by a thin gaseous envelope at most. Though these solid planets may cool over time, this cooling will not impact their radii significantly. For sub-Neptunes and more massive planets, cooling may impact the planet size significantly (e.g. Vazan et al. 2018). This essentially means that for super-Earth planets and smaller, the
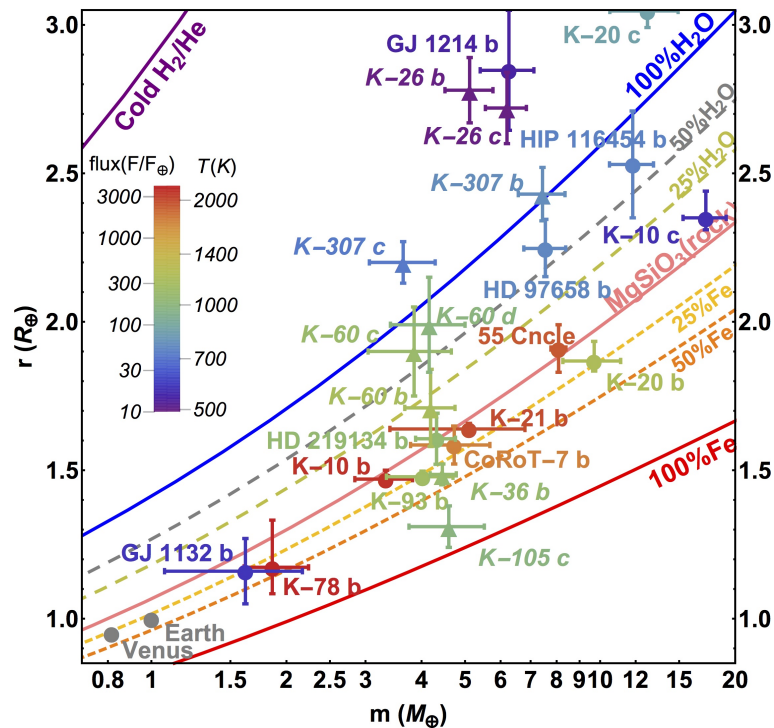
Figure 12.12: *Mass-radius diagram of planets up to 20 earth masses with better than 30% mass measurement accuracy as of 10/2016. Planets with masses measured by transit timing variation (TTV) are shown as triangles, and planets with masses measured by ground-based radial velocity are shown as circles. Curves show models of different compositions, with solid indicating single composition (Fe, MgSiO3, i.e. rock, $H_2O$) and dashed indicating Mg-silicate planets with different amounts of $H_2O$ or Fe added. Planets are color-coded by their incident bolometric stellar flux (compared to the Earth) and equilibrium temperatures. Equilibrium temperatures and fluxes are calculated assuming circular orbits, uniform surface temperature (i.e. complete heat distribution over the planetary surface), and zero-albedo. From: L. Zeng et al. 2016, ApJ 819, 127 (see: https://www.cfa.harvard.edu/ lzeng/planetmodels.html#mrrelation)*

present-day $M$-$R$ relation reflects the outcome of the planet formation process, which is why we spend a final section on this topic.

Planets with both mass and radius measurements can be plotted on a mass-radius diagram. These two physical properties are helpful in their characterisation as they provide clues about the planetary composition. Traditionally, planets have been divided into two main groups. The first consists of relatively small, terrestrial planets composed of refractory materials while the second is that of the gas-dominated planets (e.g. Weidenschilling 1977). In part, this division is inspired by the Solar System. However, the diversity in masses and radii of exoplanets discovered over the years has motivated the introduction of new planet types, including sub-Earth and super-Earth planets, and sub-Neptunes.

These planet types represent a range of planet masses and internal structures. Probably most planets are differentiated, i.e. consist of a core and one or more surrounding shells. The relative sizes of the core and shell(s) may differ, even for the same planet mass. Model curves for different internal structures on the mass-radius diagram are key to understanding the internal composition and structure of all these planets.

One may derive the internal structure of planets by combining three equations: the mass of a spherical shell

$$\frac{\mathrm{d}m(r)}{\mathrm{d}r} = 4\pi r^2 \, \rho(r), \tag{12.27}$$

the equation of hydrostatic equilibrium (Eq. 5.1)

$$\frac{\mathrm{d}p(r)}{\mathrm{d}r} = -\frac{G\,m(r)\,\rho(r)}{r^2} \tag{12.28}$$

and the EOS (Eq. 3.1)

$$p(r) = f(\rho(r), T(r)), \tag{12.29}$$

where $f$ is a unique function for a given material. The equations (12.27) and (12.28) are numerically integrated starting at the planet's center ($r = 0$) using the inner boundary condition $M(0) = 0$ and $P(0) = P_{\mathrm{central}}$, where $P_{\mathrm{central}}$ is a chosen central pressure. For the outer boundary condition $P(R) = 0$, where $R$ is the planet radius. Integrating these equations over and over for a range of $P_{\mathrm{central}}$ provides the mass-radius relationship for a given EOS.

For differentiated planets containing two or more kinds of material, one specifies the desired fractional mass of the core and of each shell. One then integrates Eqs. (12.27) and (12.28) as specified above, given a $P_{\mathrm{central}}$ and outer boundary condition, and switches from one material to the next (hence one EOS to the next) where the desired fractional mass is reached, using a guess of the total planet mass. Since the total mass that a given integration will yield is unknown ahead of time, one generally needs to iterate a few times in order to produce a model with the desired distribution of material.

Seager et al. (2007, ApJ 669, 1279) summarize equations of state that assume uniform or zero temperature for a set of materials and material conditions. For solid exoplanets, temperature has little effect on the radius. The reason is that the density of a solid changes by a relatively small amount under the influence of thermal pressure. At relatively low pressures, the crystal lattice structure dominates the material's density and the thermal vibration contribution to the density is small in comparison. At relatively high pressures, the close-packed nature of the atoms prevents any significant structural changes from thermal pressure contributions.

For pressures below approximately 200 GPa[2] Seager et al. adopt the common EOS formulae of either Vinet (1987, 1989) or Birch and Murnagham (BM equation of state or BME). For a derivation of these EOS see Poirier (2000). The Vinet EOS is

$$P = 3K_0 \, \eta^{2/3} \left( 1 - \eta^{-1/3} \right) \exp\left[ \frac{3}{2} \left( K_0' - 1 \right) \left( 1 - \eta^{-1/3} \right) \right]. \tag{12.30}$$

---

[2]For reference, the pressure at the center of the Earth is about 365 GPa.

| Atom or compound | $K_0$ | $K_0'$ | $K_0''$ | $\rho_0$ | EOS |
|---|---|---|---|---|---|
| | GPa | | GPa$^{-1}$ | Mg m$^{-3}$ | |
| C (graphite) | $33.8 \pm 3$ | $8.9 \pm 1.0$ | | 2.25 | BME3 |
| Fe ($\alpha$) | $162.5 \pm 5$ | $5.5 \pm 0.8$ | | 7.86 | BME3 |
| Fe ($\epsilon$) | $156.2 \pm 1.8$ | $6.08 \pm 0.12$ | | 8.30 | Vinet |
| FeS | $35 \pm 4$ | $5 \pm 2$ | | 4.77 | BME3 |
| H$_2$O (ice VII) | $23.7 \pm 0.9$ | $4.15 \pm 0.07$ | | 1.46 | BME3 |
| H$_2$O (liquid) | 2.28 | | | | |
| MgO | $177.0 \pm 4$ | $4.0 \pm 0.1$ | | 3.56 | BME3 |
| MgSiO$_3$ (enstatite) | 125 | 5 | | 3.22 | BME3 |
| MgSiO$_3$ (perovskite) | $247 \pm 4$ | $3.97 \pm 0.04$ | -0.016 | 4.10 | BME4 |
| (Mg$_{0.88}$, Fe$_{0.12}$)SiO$_3$ (perovskite) | $266 \pm 3$ | $4.1 \pm 0.1$ | | 4.26 | BME3 |
| SiC | $227 \pm 3$ | $4.1 \pm 0.1$ | | 3.22 | BME3 |

Table 12.3: *Parameters for the Vinet and Birch-Murnagham (BME) EOS fits. The phase of the measured material is given within brackets in the first column. Data compiled by Seager et al. (2007, ApJ 669, 1279).*

Here $\eta = \rho/\rho_0$ is the ratio of the actual density to the density of the material at room temperature and average pressure at mean sea-level on Earth (which is $1.013 \times 10^5$ Pa).

The third-order finite strain BME is given by

$$P = \frac{3}{2} K_0 \left( \eta^{7/3} - \eta^{5/3} \right) \left[ 1 + \frac{3}{4} \left( K_0' - 4 \right) \left( \eta^{2/3} - 1 \right) \right]. \tag{12.31}$$

For the fourth-order finite strain BME, the term

$$+ \frac{3}{2} K_0 \left( \eta^{7/3} - \eta^{5/3} \right) \frac{3}{8} K_0 \left( \eta^{2/3} - 1 \right)^2 \times \left[ K_0 K_0'' + K_0' \left( K_0' - 7 \right) + \frac{143}{9} \right] \tag{12.32}$$

is added to Eq. (12.31). The majority of experiments (from which $K_0$ and $K_0'$ are derived) are typically limited to pressures less than 150 GPa and temperatures less than 2000 K. Table 12.3 lists the $K_0$, $K_0'$, and $K_0''$, and the type of EOS for a number of materials. For a discussion of EOS's for higher pressures, see Seager et al. (2007).

### Exercise 12.1

a) Is our moon within the Hill sphere radius of Earth?

b) Is an astronaut working just outside of the International Space Station within the Hill sphere radius of the ISS? The mass of the ISS is 104 tonnes, the radii of the modules are at least 2 m. ISS is in orbit at a height of 300 km.

### Exercise 12.2

a) Derive an equation that establishes the radius $R_{\text{weak}}$ of the most vulnerable bodies, from Eq. (12.8).

b) Determine for each material type in Table (12.1) the radius of the weakest body. Bodies that are larger are held together primarily by self-gravity.

### Exercise 12.3

a) Asteroids in the main belt with $e \simeq 0.1$ would be expected to collide today with typical velocities of the order of $2\ \text{km}\,\text{s}^{-1}$. Adopt a mass ratio $m/M = 0.1$. Up to which target radius would this lead to catastrophic disruption?

b) Is it realistic to assume that the typical present-day random velocities that characterize collisions in the asteroid belt are also typical for the epoch of planet formation in our solar system, some 4.6 Gyr ago?

### Exercise 12.4

In the case of strong gravitational focussing, the rate of increase of mass of a proto-planet is given by $dm/dt \propto m^{\alpha}$, with $\alpha = 4/3$ (see Eq. 12.12). Derive the value of $\alpha$ for the case of negligible gravitational focussing.

### Exercise 12.5

Write a computer program that solves Eqs. (12.27) and (12.28) for a planet composed of perovskite. Apply for the equation of state the third-order strain BME and the constants $K_0$ and $K_0'$ given in Table 12.3. Compute the mass and radius for values of the central pressure $P_{\text{central}}$ of 50, 100, 150, 200, 250, and 300 GPa and draw the $M$-$R$ points in Fig. 12.12.